

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

DM protein - protein search, using sw model

Run on: October 28, 2003, 15:46:51 ; Search time 86 Seconds

(without alignments)
1002.192 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Perfect score: 2901

Sequence: 1 NAEGETAVPVVPGYHKNEI.....IINTLKGGTDGNTGHRNL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A Geneseq 19Jun03.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length DB ID	Description
1	2901	100.0	618 22	Lyophyllum shimeji
2	1142.5	39.4	623 20	Coriolus versicolor
3	1142.5	39.4	623 23	Pyrenopeziza p
4	1136	39.2	633 21	Pleurotus cornucop
5	1128.5	38.9	623 17	Pyrenopeziza p
6	1119	38.6	622 22	Trametes hirsuta p
7	1024.5	35.3	566 20	An antitumour prot
8	976.5	33.7	564 21	T. matsutake pyran
9	976.5	33.7	564 21	Trichoderma derive

10	192	6.6	51	20	AAW87541	Peptide derived fr
11	177	6.1	67	21	AAE10459	T. matsutake pyran
12	177	6.1	67	21	AAE10459	N-terminal fragmen
13	173.5	6.0	539	23	AAU97826	Glucose dehydrogen
14	170	5.9	1209	22	AAU97826	2-keto-D-gluconate
15	168	5.8	34	22	AAU97826	Lyophyllum shimeji
16	162	5.6	30	22	AAU97826	Lyophyllum shimeji
17	159	5.5	615	20	AAU97826	Glucuronate dehydrog
18	157	5.4	544	20	AAU97826	G. oxydans D-sorbi
19	142	4.9	29	22	AAU97826	Lyophyllum shimeji
20	140.5	4.8	562	23	ABB09452	Glucose-3-dehydrog
21	130.5	4.5	58	20	AAW87540	Peptide derived fr
22	128.5	4.4	657	21	AAU97826	Arabidopsis thalia
23	128.5	4.4	675	21	AAU97826	Arabidopsis thalia
24	128.5	4.4	748	21	AAU97826	Arabidopsis thalia
25	121	4.2	24	22	AAU97826	Lyophyllum shimeji
26	115	4.0	50	20	AAW87544	Peptide derived fr
27	114.5	3.9	704	20	AAU97826	Herbicidally activ
28	114	3.9	1012	23	ABB09452	Streptomyces globi
29	110	3.8	551	21	AAU97826	Listeria monocytog
30	110	3.8	1498	23	ABB09452	Peptide derived fr
31	108	3.7	35	20	AAW87544	Chlamydia pneumoni
32	106.5	3.7	775	23	ABB09452	Chlamydia pneumoni
33	106.5	3.7	778	20	AAU97826	Choline oxidase (C
34	106	3.7	546	20	AAU97826	Protonibacterium
35	105	3.6	609	22	AAU97826	Drosophila melanog
36	105	3.6	833	22	ABB09452	Drosophila melanog
37	104.5	3.6	1356	22	ABB09452	Phosphatidyl inosi
38	104.5	3.6	1876	18	AAW8757	Drosophila melanog
39	104.5	3.6	1876	22	ABB09452	Drosophila melanog
40	104.5	3.6	1876	22	ABB09452	Novel human diagno
41	103	3.6	437	22	ABB09452	Drosophila melanog
42	103	3.6	1688	22	ABB09452	Peptide derived fr
43	102.5	3.5	61	20	AAW87535	Drosophila melanog
44	102.5	3.5	352	22	ABB09452	Human class II PI3
45	102.5	3.5	1686	19	AAW09991	

ALIGNMENTS

RESULT 1
AAE97035
ID AAE97035 standard; Protein; 618 AA.

XX AAE97035;

AC AAE97035;

DT 20-JUL-2001 (first entry)

XX Lyophyllum shimeji antibacterial protein.

DE Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;

KW Pyricularia oryzae; Rhizoctonia solani; rice pathogen.

XX Lyophyllum shimeji.

OS WO200121657-A1

PN 29-MAR-2001

PD 20-SEP-2000; 2000WO-JP06404.

XX 21-SEP-1999; 99JP-0267238.

XX (NISR) JAPAN TOBACCO INC.

PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.

XX Takakura Y, Kuwata S, Inoue Y;

XX WPI; 2001-281598/29.

XX N-PSDB; AAF99980.

XX Antibacterial protein and encoded gene isolated from Lyophyllum

KW	Coriolus versicolor; modified; pyranose oxidase; heat stability;
KW	pH stability; glucose; glucosone.
XX	
OS	Coriolus versicolor.
XX	
PN	JF11009271-A.
XX	
PD	19-JAN-1999.
XX	
PF	24-JUN-1997; 97JP-0181865.
XX	
PR	24-JUN-1997; 97JP-0181865.
XX	
PA	(KIKK) KIKKOMAN CORP.
XX	
DR	WPI; 1999-145891/13.
XX	
PT	New modified pyranose oxidase gene and protein - useful for
PT	efficient recombinant production of pyranose oxidase with excellent
PT	Km, heat and pH stability
XX	
PS	Claim 1; Page 7-9; 10pp; Japanese.
XX	
CC	The present sequence represents pyranose oxidase derived from Coriolus
CC	versicolor, which can be modified by adding, deleting or replacing at
CC	least one amino acid [preferably where the amino acid at position 542
CC	is replaced] and still retains pyranose oxidase activity. The modified
CC	pyranose oxidase with the amino acid at position 542 replaced, has the
CC	following properties: (a) Action: It oxidises glucose to glucosone;
CC	(b) Substrate specificity: It reacts specifically with glucose and also
CC	reacts with galactose, L-sorbose, D-xylose and 1,5-anhydro-D-glucitol;
CC	(c) Stable pH: 3.5 to 11.0 at 50 degrees Celsius for 30 minutes; (d)
CC	Optimum pH: Near 6.5; (e) Optimum temperature: Near 55 degrees Celsius;
CC	and (f) Temperature stability: Stable up to 55 degrees Celsius. The
CC	present invention also describes a method for the preparation of a
CC	modified pyranose oxidase. The method efficiently prepares a modified
CC	pyranose oxidase with excellent Km value, heat stability and
CC	pH stability.

		Query Match	39.4%; Score 1142.5; DB 20; Length 623;
		Best Local Similarity	45.1%; Pred. No. 5.4e-99;
		Matches	251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;
QY	13	G Y H K N E I F Q K D I R F V N V I K A L C O Q V S V P V R N Q N V P T L D P G A W S A P P G S S A I S N G K I P	72
DB	87	G A H K N T V E Y Q K N I D F K V N V I Q G L M S V S V P V N T L V I D L S P T S W Q A - - S S F P V R N G S N P	144
QY	73	H Q R E F E N L S A E A V T R O V G M G M S T H W T C S T P R I H P P M E S L P G I G R P K U S N D P A E - D D K E W N E	131
DB	145	E Q D P L E N L S G Q A V T R V V G M G M S T H W T C A T P R F O R E Q - - - - - R P L L V K D Q D D A D A E W D R	197
QY	132	L Y X E A E R L G T S K F D E S I R H T L V L R S O D A Y K D R O I F R P L P L A C H R L K N K A P E Y V E W H	191
DB	198	L Y T K A S Y K T G D Q P K E S I R N L N L V N K L A E Y K - Q R D F Q Q I P L A A T R - - R G P T F V E W S	254
QY	192	S A E N L P - H S I Y N D D R Q K K L F T L T N H R C T R L A L T G G Y E K K I G A A E V R N L L A T R N P S S Q L	249
DB	255	S A N T V F D L Q N R P N T D A P N E R F N L F P A V A C E R V - - - - - V R N - - - - - T S N S E I	295
QY	250	D S Y - - - - - I M A K V Y V L A S A I G N P Q I L Y N S G F S G L - Q V T P R N - D S L I P N L G R Y	295
DB	296	E S L H I H D L I S G R F E I K A D V F L T A G A V N A Q L L V N S G F G Q L G R P D P A N P Q P L P L S G S Y	355
QY	296	I T E Q P M A F C Q I V L R Q B F V D S V R D D P Y - - G L P - - - - - M W K E A V A Q H	333
DB	356	I T E Q S L V F C Q T W S I T E L I D S V K S D M I I R N G P D L G Y S V I T Y P C A E T N K H P D W N E K V Q N H	415
QY	334	I A N Q P T D A L P I P F R D P E Q V T P T F T E H P W H Q I O H R D A F S Y G A V G P E V S R V I D L W R F G	393
DB	416	M N M Q O D P L I P F E D P E Q V T T L F O P S H P M W T Q I H R D A F S Y G A V Q A Q S I D S R L V D W R F F G	475

Db	296	ESLHIHDLISGRFEIKADVFVLTAGAVHNAQLVNSGFGQLGRDPANPPQLPLSLGSY	355
Qy	296	ITEQPMAFQIQLRQEFVDSVRDDPY--GLP-----WWKEAAVQH	333
Db	356	ITEQSLVFCQTVNSTELIISVKSDMIIRGNPGDLGYSVVTTPGAETKNKPDWNEKVKVNH	415
Qy	334	IAKNPTDALIPFRDPEPQVTTFTTEBHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFG	393
Db	416	MMQHQEDPLPIPEDPEPQVTTTLFQSPHPWHTQIHRDAFSYGAVQQSIDSRLIYDWRFFG	475

QY 394 ATDPENALLVQNDVQGVSMPOPTFRYR--PSTASNVARKMMADMCCEVASNLGGYLPT 452
 Db 476 RTEPKEENKLFWSKIDTDTYNMPOPTDFRFPAGRTSKEAEDMMTDMCVMSAKIGGLPG 535
 QY 453 SPQFMDPGLALHLAGTTRIGFDKA--TTVADNNSLVDFANLVVAGNTRIGFGENPT 510
 Db 536 SLQFMEFGLVHLGGTHRGFDQEDKCCVNTDSRVFGFNLFGLGCGNIPTAYGANPT 595
 QY 511 LTSMCHAIKSARSIIIN 526
 Db 596 LTAMSLAIKSCYIKN 611

RESULT 3

AAO15842
 IID AAO15842 standard; Protein; 623 AA.

XX AC AAO15842;

DT 19-DEC-2002 (first entry)

XX

Pyranose oxidase protein.

Recombinant polypeptide display method; signal peptide; pyranose oxidase;
 transmembrane linker; transporter domain; autotransporter; enzyme;
 protease recognition site; catalytic biomolecule design;
 cell factory design; adrenotoxin; mitochondrial target sequence.

XX Unidentified.

XX WO200270645-A2.

XX 12-SEP-2002.

XX 01-MAR-2002; 2002WO-BP02246.

XX 02-MAR-2001; 2001EP-0105129.

XX (PLAC) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.

XX Jose J, Hannemann F, Bernhardt R;

XX WPI; 2002-732791/79.

XX Displaying recombinant functional polypeptide containing prosthetic
 group or several subunits, on host cell surface using the transporter
 domain of an autotransporter which is fused to a signal peptide -

XX Disclosure; Fig 14; 80pp; English.

XX The invention comprises a method of displaying a recombinant polypeptide
 containing a prosthetic group on the surface on the host cell. The method
 involves transforming a host cell with a nucleic acid construct that
 encodes a signal peptide, the polypeptide to be displayed, a
 transmembrane linker, a transporter domain of an autotransporter and
 optionally a protease recognition site. The method of the invention is
 useful for displaying a recombinant polypeptide containing a prosthetic
 group on the surface of a host cell, where the prosthetic group comprises
 an inorganic component such as a metal (e.g. cobalt, nickel, manganese,
 copper or iron) containing group. The method of the invention has
 applications in evolutive design of catalytic biomolecules or for new
 whole cell factories. The present amino acid sequence represents a
 C pyranose oxidase protein.

XX Sequence 623 AA;

XX Query Match

XX Best Local Similarity 39.4%; Score 1142.5; DB 23; Length 623;

XX Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;

Y

13 GYHKNETFQKIDRFNVNKGALQQVSVPRNVNPTLDPGAWSAPPGSSAISNGKNP 72

b

87 GAHKNTVEYQKIDKFVNVIQQLMSVSVFNTLVIDTLSTSQQA--SSFFVRNGSNP 144

QY 73 HOREFENLSAEAVTRGVGGMSTHTCTSTPRIHPHBMESLPGIGRPKLSNDPAE--DDKEWNE 131
 Db 145 EQDLNLSQAVTRVVGNMSTHTCATPRFDEQ-----RPLLVKDDQDADDAEWDR 197
 QY 132 LYSEAEERLTGTSKBFDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEYVEWH 191
 Db 198 LYTAKSSYFKTGTDQFKESIRHNLVLNKLAEYKG--QRDFQOIPLAATR--RSPTTFVWS 254
 QY 192 SAENLF--HSIYNDDKQKLLFTLLTHRCTRLALTGGYEKKIGAAEVRNLLATRNPSQL 249
 Db 255 SANTVFELQNRPNTPADNERFNLFPVACERV-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKYVVLASCAIGNPQILYNSGFSGL--QVTPRN--DSLIPNLGRY 295
 Db 296 ESLHIHDLISGRFEIKADVFLTAGAVHNAQLVNSGFGQLGRPDPAFPQLPLSLGSY 355
 QY 296 ITEQPMAFQOIVLRQEFVDSVRDDPY--GLP-----MWKEAVAQH 333
 Db 356 ITEQSLVFCQVMSTELIDSVKSDMIIRGNPGDLGYSVTYTPCAETNKHDPWNNEKVXH 415
 QY 334 IAKNPTDALPIPRDEPOVTPPTEEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWF 393
 Db 416 MMQHOEDPLPIPFEDPEPQVTTLPQSPHFWHTQIHRDAFSYGAVQOSIDSLIVDMRFFG 475
 QY 394 ATDPEANLLVFQNDVQGVSMPOPTFRYR--PSTASNVARKMMADMCCEVASNLGGYLPT 452
 Db 476 RTEPKEENKLFWSKIDTDTYNMPOPTDFRFPAGRTSKEAEDMMTDMCVMSAKIGGLPG 535
 QY 453 SPQFMDPGLALHLAGTTRIGFDKA--TTVADNNSLVDFANLVVAGNTRIGFGENPT 510
 Db 536 SLQFMEFGLVHLGGTHRGFDQEDKCCVNTDSRVFGFNLFGLGCGNIPTAYGANPT 595
 QY 511 LTSMCHAIKSARSIIIN 526
 Db 596 LTAMSLAIKSCYIKN 611

RESULT 4

AAV52700
 ID AAV52700 standard; Protein; 633 AA.

XX AC AAV52700;

XX DT 07-MAR-2000 (first entry)

XX

DE Pleurotus cornucopiae antitumour protein.

XX

KW Antitumour; cancer; tumour; treatment; expression; tumour suppressor;

XX P53; pBR; ss.

XX OS Pleurotus cornucopiae.

XX

PN JP11315096-A.

XX PD 16-NOV-1999.

XX PF 07-AUG-1998; 98JP-0216349.

XX

XX PR 08-AUG-1997; 97JP-0215311.

XX

XX PR 02-MAR-1998; 98JP-0066176.

XX

XX (NEWF-) NEW FOOD CREATION GIJUTSU KENKYU KUMIAI.

XX PA

XX WPI; 2000-058170/05.

XX

XX DR N-PSDB; AA46411.

XX

PT An antitumour protein derived from Pleurotus cornucopiae and its gene
 PT useful for treatment of cancer including those caused by abnormal
 PT expression of cancer inhibitory gene (e.g. p53 and pBR)

XX

PS Claim 3; Page 13-15; 23pp; Japanese.

CC The invention relates to a novel antitumour protein extracted from
 CC fruiting bodies of the fungus *Pleurotus cornuopae*. The protein and
 CC nucleotides encoding it are useful for the treatment of cancer,
 CC including those caused by abnormal expression of tumour suppressor
 CC genes such as p53 and pBR. This sequence represents the antitumour
 CC protein.

XX SQ Sequence 633 AA;

Query Match 39.2%; Score 1136; DB 21; Length 633;
 Best Local Similarity 45.3%; Pred. No. 2.3e-98;
 Matches 248; Conservative 70; Mismatches 175; Indels 54; Gaps 15;

QY 13 GYHKNEIEFQKIDRFVNVIKALQOVSVPRNQNVPTLDPGAWAPPSSAISNGKNP 72
 DB 96 GSHKNTVEYQKNDKFNVIQQLMPVSVPRNKYVADTLSPASWOA--STIFVRNGANP 153
 QY 73 HOREPENLSAEAVTRGVGGMSTHTCTSTPRIHPPMESLPGIGRPKL-SNDPAEDDKWNE 131
 DB 154 EQDFTNLSCQAVTRVVGGMATHTWTATPRFHS-----EPKLVKDDSDADALEWER 206
 QY 132 LYSEAEERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEYVEWH 191
 DB 207 LYDIAESFVKTGHQFQDSIRHNLVLEKLSYAG-QRGFEQIPLAAQRTN--PRFVENS 263
 QY 192 SAENLF--HSIYNDKQKLLFTLLTNHRCRTRALTGGEYKKGAAAEVRNLLATRNPSQL 249
 DB 264 SAHTVFDLENRPNADDEKGRNLFPAVVCERVTRDSDRKIDIEDIEVHDLISG----- 315
 QY 250 DSI-IMAKVYVLASGAIGNPQILYNSGSGLVTPRND---SLIPNLGRYTEQPMAF 304
 DB 316 DRYKXADVFILCSGAVHNQILVNSGF-GRMGQDSSLPPTLLPYLSYITEQTLTFC 374
 QY 305 QIVLRQEFVDSVRDDP--YGLP-----WKEAQAQHIKXNPTDAL 342
 DB 375 QTVFSTELNVLKSDMIIVGTPGPDYVTFPTDPSNKHNNWNEKVKQHMHOEDPL 434
 QY 343 PIFRDEPEQVTPFTTEHHPWHTQIHRDAFSYGAVGSDRSRVIVDLRMFGATDPANLL 402
 DB 435 PFLDDPEQVTTLFQDTHPWHQIHRDAFSYGAVAESIDSRILVDMRFFGRTEPVEENK 494
 QY 403 LVQNDVQDGYSMQPTFRYR--PSTASNVRAKWMADMCVAGNLGGLYPTSPQWMDPG 461
 DB 495 LWFESKQITDAYNLPQPTFSRFFQGRTAQAEELMADMCMTNMTKVGGLFGSYPQWMAFG 554
 QY 462 LALHLACTTRIGFDKATTV--DNNSLVMDFANLYVAGNGTIRTFGEENPTLTSMCHAIK 519
 DB 555 LVLHLGCTHRMGFDEADKACVDNTSKVFCMENFLGCGCGNICTAYASNPTLTALVALAIR 614
 QY 520 SARSIIN 526
 DB 615 SKYIRN 621

RESULT 5

AAR99628 standard; Protein; 623 AA.

XX AAR99628;

XX AAR99628;

DT 27-NOV-1996 (first entry)

XX Pyranose oxidase.

XX Pyranose oxidase; glucose; oxidation; glucanase; assay; diabetes;
 KW marker; diagnosis; 1,5-anhydro-D-sorbitol.

XX Coriolus versicolor.

XX DE19545780-A1.

XX 13-JUN-1996.

PF 07-DEC-1995; 95DE-1045780.
 XX 24-MAY-1995; 95JP-0124835.
 PR 07-DEC-1994; 94JP-0304086.
 XX (KIKK) KIKKOWAN CORP.
 XX Kawai G, Koyama Y, Minamihara T, Nishimura I, Okada K;
 PI Suzuki M;
 XX WPI; 1996-278990/29.
 DR N-PSDB; AAT34420.
 XX DNA encoding protein with pyranose oxidase activity at neutral pH -
 PT useful for the determination of glucose in body fluids or foods, or
 PT 1,5-anhydro-D-sorbitol used as marker for diabetes diagnosis
 XX Claim 1; Page 14-16; 22pp; German.
 XX The present sequence encodes a protein isolated from *Coriolus versicolor*,
 CC which has the enzyme activity of pyranose oxidase (PO). The PO oxidises
 CC glucose to gluconone and has an optimum pH of 7-7.5. It has a mol. wt. of
 CC 29000 (determined by gel filtration) and is stable at around 50deg.C.
 CC PO can be used for measurement of glucose in, e.g. foods or body fluids,
 CC or 1,5-anhydro-D-sorbitol which is an important marker used in the
 CC diagnosis of diabetes.
 XX SQ Sequence 623 AA;
 Query Match 38.9%; Score 1128.5; DB 17; Length 623;
 Best Local Similarity 45.0%; Pred. No. 1.1e-97;
 Matches 250; Conservative 68; Mismatches 165; Indels 73; Gaps 15;
 QY 13 GYHKNEIEFQKIDRFVNVIKALQOVSVPRNQNVPTLDPGAWAPPSSAISNGKNP 72
 DB 87 GAKKNTVEYQKNDKFNVIQQLMSVSPVNTLVITLSPTSWOA--SSFFVRNGSNP 144
 QY 73 HOREPENLSAEAVTRGVGGMSTHTCTSTPRIHPPMESLPGIGRPKLSNDPAE--DDKEWNE 131
 DB 145 EQDFTNLSCQAVTRVVGGMSTHTWTATPRFREQ-----RPLVKDDQDADDAEWDR 197
 QY 132 LYSEAEERLIGTSTKEFDESIRHTLVLSLQDAYKDRQIRFRPLPLACHRLKNAPEYVEWH 191
 DB 198 LYTAESEYFKTQDQPKESIRHNLVNLKLAEEYKG-QRDFQIPLAATR--RSPTTFVEMS 254
 QY 192 SAENLF--HSIYNDKQKLLFTLLTNHRCRTRALTGGEYKKGAAAEVRNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPADNERFNLFPVAVACERV-----VRN-----TSNSEI 295
 QY 250 DSI-----IMAKVYVLASGAIGNPQILYNSGSGLV--QVTPRN--DSLIPNLGRY 295
 DB 296 ESLHIHDLISGDRFEIKADVFLITAGAVHNAQLLVNSGFGQLGRPDNPANPQLPLSURY 355
 QY 296 ITEQPMAFQIVLRQEFVDSVRDDPY--GLP-----WKEAQAQHI 333
 DB 356 ITEQSLVFCQTVMSSTELIDSVKSDMIIRGNPGDLGYSVTVTPCAETNKHFDWNEKVKNH 415
 QY 334 IAKNPTDALPIPRDPEQVTPFTTEHHPWHTQIHRDAFSYGAVGSDRSRVIVDLRMFG 393
 DB 416 MMQHQEDPLPIPRDPEQVTPFTLQPSHPWHTQIHRDAFSYGAVQVQITDRLIVDMRFFG 475
 QY 394 ATDPEANLLVFQNDVQDGYSMQPTFRYR--PSTASNVRAKWMADMCVAGNLGGLYPT 452
 DB 476 RTEPKEENKLFSDKITDTYNMNPQPTDFRFPAGRTSKAEADMMTDMCVNWKIGGLPLG 535
 QY 453 SPQFMDPGLALHLACTTRIGFDKA--TTVADNNSLVMDFANLYVAGNGTIRTFGEENPT 510
 DB 536 SLQPFMEPGLVHLGCTHRMGFDEQEDKCCVNTDSRVFGKFLNLFGLGCGNIPRAYGANPT 595
 QY 511 LTSMCHAIKSARSIIN 526
 DB 596 LTAMSLAIKSEYIKN 611

	Query Match	35.3%	Score 1024.5	DB 20	Length 566
Best Local Similarity	43.0%	Pred. No. 7.3e-88			
Matches 232	Conservative 77	Mismatches 180	Indels 51	Gaps 12	
QY	2	AEEGTAVPYGVYHKHKNIEFQKQIDRFNVNVIKGALQOVSPVRNQNVPTLDPGASAPP	61		
DB	62	AEIGSQDNPFVGAHHKNSIKFKQDKDFWNIINGALOPISISPSDTYPTFLVAAM-APE	120		

Qy	62	GSSA-----LSGNKPHQREFENLSAEAVTRGVGGGNSTHWTCSTPRIHPHMESLPGIGRKP	117
Db	121	IDPAEQOLVIMGHNPENQEAGLNLPGSVAVTRTVGGNATHWTCACPTPHD-----EB	170
Qy	118	LSNDPAEDDKENNELYSAEERLIGSTKEPFDSIRHTLVLSQDAYKQRIFRFLPLA	177
Db	171	RVNPNV-DKQEFALLERAKTLNVLHVSQIQYDDSIQIVVKETLQOQTL-DASRGVTTPLG	228
Qy	178	CHRLKNAPSYVEWHSAENLFHSIYNDDKQKKLFTLLTNHRCRTIALTGGYEKKIGAAEVR	237
Db	229	VERRTDNPIYVWTGAD---TVLGDVPKSPRFALVTETRVTKLIVSETNPTQVVAALLR	284
Qy	238	NLLATRNPSQLDSYIMAKYVYLASCAIGNPQILYSGFSGLOVTPRNDSLIPNLGRYT	297
Db	285	NL-----NTSN---DELWAKSFVIACGACTPQILWNS-----NIRPY-----ALGRYLS	327
Qy	298	EQMAPFCQILVROEYFVSDVRDDPYGLPWKAEVAQHIAKNPTDALPIPRDPEPQVTTPF	357
Db	328	EQSMTFQVILKRGIVDAIATDPR-----FAAKVEAHKKHPPDDVLPIPFHEPEPQVMIPY	383
Qy	358	TEHPMHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEANLLVF-----QND	408
Db	384	TSDFPMHVQVHRDAFSYGDVGPKADPRVYVDLRFFGKSDIVEENRVTFGNPKLREWEAG	443
Qy	409	VODGYSMPQPTFRPPTASVNRARKMADMCCEVASNLGGYLPSPQPMDFGLALHLHAG	468
Db	444	VTDTYGMPQPTFHVKRTNADGDRQRMNDMTNVANMLGGYLPSPQPMDFGLVLIHITG	503
Qy	469	TTRIGFDKATTVADNNSLVMPDFANLYVAGNGTIRTFGENPTLTSMCHAIKSARSIINTL	528
Db	504	TTRIGTDQTSVADTSTKVHNFNLWVGNGCIPDPTACNPTRTSVAYALKGAEAVVNYL	563
.			
RESULT	8		
ID	AAB10457		
AA	AAB10457 standard; Protein; 564 AA.		
AC	AAB10457;		
XX			
XX	11-DEC-2000 (first entry)		
DT			
XX			
DE	T. matsutake pyranose oxidase protein.		
XX			
KW	Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;		
XX	diabetes mellitus.		
XX			
XX	Tricholoma matsutake.		
OS			
XX	JP2000175698-A.		
PN			
XX			
PD	27-JUN-2000.		
XX			
PF	16-DEC-1998; 98JP-0357423.		
XX			
PR	16-DEC-1998; 98JP-0357423.		
XX			
XX	(NISB) JAPAN TOBACCO INC.		
XX			
DR	WPI ; 2000-478485/42.		
DR	N-PSDB : AAA71487.		

[illegible]

XX OS Trichoderma matsutake.
 XX XX WO200014242-A1.
 XX PN 16-MAR-2000.
 XX PD 19-AUG-1999; 99WO-JP04441.
 XX XX 08-SEP-1998; 98JP-0270606.
 XX PR (NISR) JAPAN TOBACCO INC.
 XX PA (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 XX XX Takakura Y, Kuwata S, Ohta S;
 DR WPI; 2000-256990/22.
 DR N-PSDB; AAA07403.
 XX XX Mushroom-derived antibacterial protein against plant pathogenic fungi
 PT of rice, with activity and thermal stability, obtainable cheaply on
 PT large scale, useful in agriculture
 XX Claim 3; Page 41-45; 52pp; Japanese.
 XX PS This sequence is the Trichoderma matsutake antibacterial protein of
 XX CC the invention. The protein has activity against at least Pyricularia
 XX CC oryzae and Rhizoctonia solani, and is obtained from a fraction of an
 XX CC aqueous extract of a mushroom precipitated by the ammonium sulphate
 XX CC precipitation method. The protein has a molecular weight of about 210 kD
 XX CC as determined by the gel filtration method, includes components of about
 XX CC 15 kD and 50 kD in SDS-PAGE, and is stable to heating in an aqueous
 XX CC neutral solution at 60 degrees C for 10 minutes but with loss of
 XX CC antibacterial activity after heating in the solution at 80 degrees C for
 XX CC 10 minutes. The protein is used for inhibiting the growth of plant
 XX CC pathogenic fungi e.g. Pyricularia oryzae and Rhizoctonia solani.
 XX CC It is useful in treating rice plants, and is applicable in agriculture as
 XX CC an antibacterial agent. The protein has activity at relatively low
 XX CC concentrations, and can be produced at low cost on large scale.
 XX Q Sequence 564 AA;
 Query Match 33.7%; Score 976.5; DB 21; Length 564;
 Best Local Similarity 41.3%; Pred. No. 2.6e-83;
 Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;
 Y 2 AREGTAVPVPGYHKNEIEFQKIDRFVNVIKGALQOVSVPRNQNVPTLPGASAPP 61
 b 62 ABIGSQDNPVIGAHNRNSKFKDIDKFNIIINGALQPIISPSDYQYTLAAVAW-APP 120
 Y 62 GSSA-----ISNGKNPHQREFENLSAEAVTRGVGGMSTHTWCSTPRIHPMESLPGIGRPK 117
 b 121 IDPABGQLVIMGNHNPQEBAGNLPGSAVTRTVGGMATHWCACPTPHD-----EE 170
 Y 118 LSNDPAEDDKENWELYSERLIGTSTKEPDSIRHTLVLSLQDAYKDRQRIAPLPIA 177
 b 171 RVNNEPV-DKQEFALLERAKTLNLSHQYDSDSIRQIVVVKETLQQTLDASRGVTTPLG 228
 Y 178 CHRLKNAPEYVHSAENLPHSIYNDKQKLTLLTNHCTRLALTGGYEKKIGAEVR 237
 b 229 VERRTDNIYVTTGAD----TVLGDVPKSPRFVLVETRTVTKFIVSETNPTQVVAALLR 284
 Y 238 NLLATRNPSOLDYSYIMAKVYVYLSAIGNPOILYNSGFSGLQVTPRNDLSLIPNLGRVIT 297
 b 285 NL---NTSN--DELVAQSFVIAGVACTPQLWNS-----NIRPH-----ALGRVLS 327
 Y 298 EQPMAFCQIVLRQEFVSDVRDPPYGLPWKEVAOHIAKNPTDALPIFRDEPQVTPPF 357
 b 328 EQSMTFCQIVLKRISVDSIATDPR----FAAKVEAHKKKHDPDVLPIPFHEPEQVMIPY 383
 Y 358 TEHPHWTQIHRDAFSGYAGVGPEDSVRVLDLRFEGATDPENALLVF-----QND 408
 b 384 TSDFPFWQVHR--YAFGDVGKADPRVVVDLRFEGKSDIVEENRVTFGNPKLRDWEAG 441

409 VQDGYSNPQPTFRYRSTASNVARQKMDMCEVSNLGGYLPSTPQFMDPGLALHLAG 468
 442 VTDYGMPOPTFHVKRTNADGDRDQRMNDMTNVANILGGLPGSYLPQFWAPGLAQHITG 501
 469 TTRIGFDKATTVADNNSLVWDFANLVVAGNGTTRTGFGENPTLTSMCHAIKSARSINTL 528
 502 TTRIGTDDQTSVADPTSKVHNFNVLWVGNGCIPDATACTNPTRTSVAYALKGAEEVVSYL 561

RESULT 10
 AAW87541
 ID AAW87541 standard; Peptide; 51 AA.
 XX AC AAW87541;
 XX XX 26-FEB-1999 (first entry)
 XX DE Peptide derived from an antitumour protein.
 XX KW Antitumour protein; Tricholoma matsutake.
 XX OS Tricholoma matsutake.
 XX PN JP10313876-A.
 XX PD 02-DEC-1998.
 XX PF 13-FEB-1998; 98JP-0031452.
 XX PR 13-FEB-1997; 97JP-0029275.
 XX XX (MOMO-) MOMOYA KK.
 XX PA (NORQ) NORINSUISANSHO SHOKUHN SOGO.
 XX DR WPI; 1999-074153/07.
 XX PT An anti-tumour protein - prepared by culture of host cell
 XX PT transformed by vector containing base coding sequence
 XX PS Claim 5; Page 12; 15pp; Japanese.
 XX CC The present sequence represents a peptide derived from an antitumour
 XX CC protein. Nucleic acid encoding the antitumour protein is isolated
 XX CC from Tricholoma matsutake.
 XX SQ Sequence 51 AA;
 Query Match 6.6%; Score 192; DB 20; Length 51;
 Best Local Similarity 62.7%; Pred. No. 1.5e-10;
 Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;
 Y 330 VAQHTAKNPTDALPIFRDPPQVTPTEBHPWHTQIHRDAFSGYAGVPE 380
 b 1 VEAKKKHFDVLLPIPFHEPEPQVMIPYTSDFPWHVQVHRDAFSGYDVGPK 51

RESULT 11
 AAB10459
 ID AAB10459 standard; Protein; 67 AA.
 XX AC AAB10459;
 XX XX 11-DEC-2000 (first entry)
 XX DE T. matsutake pyranose oxidase protein N-terminal fragment #2.
 XX KW Pyranose oxidase; matsutake mushroom; antibacterial; sugar metabolism;
 XX KW diabetes mellitus.
 XX OS Tricholoma matsutake.
 XX PN JP2000175698-A.

XX 27-JUN-2000.
 PD 16-DEC-1998; 98JP-0357423.
 XX 16-DEC-1998; 98JP-0357423.
 PF 16-DEC-1998; 98JP-0357423.
 PR (NISR) JAPAN TOBACCO INC.
 XX WPI; 2000-478485/42.
 DR A reagent containing pyranose oxidase for the determination of pyranose
 XX Claim 3; Page 19; 23pp; Japanese.
 PT This invention describes a novel reagent containing pyranose oxidase for
 XX the determination of pyranose which can be prepared from a fraction
 CC precipitated from an aqueous extract of matsutake mushroom by ammonium
 CC sulfate precipitation and has an antibacterial activity against at least
 CC *Pyricularia oryzae* or *Thanatephorus cucumeris* and has a M.W. of ca.
 CC 210 kD by gel filtration and shows the presence of components of ca.
 CC 50 kD and ca. 15 kD by SDS-PAGE and maintains the above antibacterial
 CC activity by being heated at 60 degrees C for 10 minutes in a neutral
 CC aqueous solution and in which the above antibacterial activity is
 CC inactivated by being heated at 80 degrees C for 10 minutes in a neutral
 CC aqueous solution. The invention also describes a method for the
 CC determination of pyranose in a sample in which the above reagent for the
 CC determination of pyranose is reacted with pyranose in the sample and the
 CC hydrogen peroxide formed is reacted with an enzyme to develop a color, a
 CC method for diagnosing a disease accompanied by abnormality in sugar
 CC metabolism in which the above reagent for the determination of pyranose
 CC is reacted with pyranose in the sample and the hydrogen peroxide formed
 CC is reacted with an enzyme to develop a color, and a kit for pyranose
 CC analysis or the diagnosis of a disease accompanied by abnormality in
 CC sugar metabolism. The pyranose oxidase can be used as a diagnostic agent
 CC for diabetes mellitus. This sequence represents a *Tricholoma matsutake*
 CC pyranose oxidase protein N-terminal fragment which is described in the
 CC method of the invention.
 XX Sequence 67 AA;
 SQ Query Match 6.1%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.2e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;
 QY 409 VODGYSMPQPTFRYPSTASNVARKMADMCCEVASNLGGYLPTSPPOFMDPGLHLHLAG 468
 DB 7 VTDYGMQPTFRHVKRTNADGDRDQRMNDNTNVANMLGGYLPGSYPOFMAPGLVHLHTG 66
 QY 469 T 469
 DB 67 T 67
 RESULT 12
 ID AA81951 standard; peptide; 67 AA.
 XX AA81951;
 AC AA81951;
 XX 07-JUL-2000 (first entry)
 DT N-terminal fragment of *Trichoderma* derived antifungal protein.
 DE Antifungal protein; *Pyricularia oryzae*; *Rhizoctonia solani*; mushroom;
 KW growth inhibitor; plant pathogenic fungi; antibacterial agent;
 KW N-terminal fragment.
 XX *Trichoderma matsutake*.
 OS WO200014242-A1.
 XX 31-OCT-2000; 2000JP-0332085.
 PR 24-NOV-2000; 2000JP-0357102.
 PR 12-SEP-2001; 2001JP-0276832.

PD 16-MAR-2000.
 XX 19-AUG-1999; 99WO-JP04441.
 PF 08-SEP-1998; 98JP-0270606.
 PR (NISR) JAPAN TOBACCO INC.
 XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & FI.
 PA Takakura Y, Kuwata S, Ohta S;
 XX WPI; 2000-256990/22.
 DR Mushroom-derived antibacterial protein against plant pathogenic fungi
 XX of rice, with activity and thermal stability, obtainable cheaply on
 PT large scale, useful in agriculture
 XX Claim 2; Page 27; 52pp; Japanese.
 PS This sequence represents an N-terminal fragment of the protein of the
 CC invention. The protein is an antibacterial protein with activity against
 CC at least *Pyricularia oryzae* and *Rhizoctonia solani* is obtained from a
 CC fraction of an aqueous extract of a mushroom precipitated by the ammonium
 CC sulphate precipitation method. The protein has a molecular weight of
 CC about 210 kD as determined by the gel filtration method, includes
 CC components of about 15 kD and 50 kD in SDS-PAGE, and is stable to
 CC heating in an aqueous neutral solution at 60 degrees C for 10 minutes but
 CC with loss of antibacterial activity after heating in the solution at 80
 CC degrees C for 10 minutes. The protein is used for inhibiting the growth
 CC of plant pathogenic fungi e.g. *Pyricularia oryzae* and *Rhizoctonia solani*.
 CC It is useful in treating rice plants, and is applicable in agriculture as
 CC an antibacterial agent. The protein has activity at relatively low
 CC concentrations, and can be produced at low cost on large scale.
 XX Sequence 67 AA;
 SQ Query Match 6.1%; Score 177; DB 21; Length 67;
 Best Local Similarity 57.4%; Pred. No. 6.2e-09;
 Matches 35; Conservative 6; Mismatches 20; Indels 0; Gaps 0;
 QY 409 VODGYSMPQPTFRYPSTASNVARKMADMCCEVASNLGGYLPTSPPOFMDPGLHLHLAG 468
 DB 7 VTDYGMQPTFRHVKRTNADGDRDQRMNDNTNVANMLGGYLPGSYPOFMAPGLVHLHTG 66
 QY 469 T 469
 DB 67 T 67
 RESULT 13
 ID AAU97826 standard; Protein; 539 AA.
 XX AAU97826;
 AC AAU97826;
 XX 27-AUG-2002 (first entry)
 DT Glucose dehydrogenase associated protein #2.
 DE Glucose dehydrogenase; electrode; glucose sensor; glucose-assay;
 XX substrate specificity; heat stability.
 KW *Burkholderia cepacia*.
 OS WO200236779-A1.
 XX 10-MAY-2002.
 PD 31-OCT-2001; 2001WO-JP09556.
 PF 31-OCT-2000; 2000JP-0332085.
 PR 24-NOV-2000; 2000JP-0357102.
 PR 12-SEP-2001; 2001JP-0276832.

XX (SODE/) SODE K.
XX Sode K;
XX WPI; 2002-463413/49.
XX N-PSDB; ABK52367.
XX Production of Burkholderia glucose dehydrogenase for use in glucose
XX sensor electrodes and glucose-assay kits in medicine, science and
XX industry
XX
XX Claim 11; Page 56-57; 61pp; Japanese.
XX
XX The invention describes a method of producing a glucose dehydrogenase
XX comprising culturing a Burkholderia microorganism and collecting the
XX product from the medium and/or the microbial cells. Glucose dehydrogenase
XX is useful in electrodes of glucose sensors and glucose-assay kits for
XX medicine, science and industry. Glucose dehydrogenase is economically
XX produced with high substrate specificity and improved heat stability to
XX provide long-term accuracy. This is the amino acid sequence of a glucose
XX dehydrogenase associated protein described in the invention.
XX
XX Sequence 539 AA;

Query Match 6.0%; Score 173.5; DB 23; Length 539;
Best Local Similarity 19.9%; Pred. No. 4.3e-07;
Matches 109; Conservative 80; Mismatches 191; Indels 169; Gaps 24;
YY 60 PFSSAISNGKPHOREFENLSAEAVTRGVGMSTHTCTSTPRIHP---PMESLPGIGRP 116
YY 78 PNDVYLKGEHKSQY-----IRAVGGTTHWAASAWRFIPNDFKMSVYGVGR- 128
YY 117 KLSNDPAE--DKEMNELYSEAEELIGTSTKEFDESIRHTLVLSLQDAYKDRQIP--RP 173
YY 129 ---DWPIQYDDE--PYORAEELGVGMGPGE-----DLYSPKQPVMP 171
YY 174 LPLACHR--LKNAPVEVHSAENLPHSI-----YNDKQKLLFLLTNHRCRTRIALTG 226
YY 172 LPLSFNEOTIKTA-----LNNYDPKPHVVTPEVARNSRPYDGRPTCCGNNCWPICPIGA 226
YY 227 Y-----EKKIGAAEVRNLLATR---NPSSQL-----DSYIMAKVYVLASG 263
YY 227 MYNGIVHVEKAERAGAKLIENAVVTKLETPDKRIVAAALYKDTGAHRVVEGYFVLAAN 286
YY 264 AIGNPQILYNSGFGSLQVTPRNDSLIPNLGRYITEQPMACQIVLRQEPVDSVDDPYGL 323
YY 287 GIETPKIL-----LMSANRDFPNGV 306
YY 324 PWKEAVAOHIAKNPTDALPI-----PFRDPEQVTPPTTEHPHMT-----QIH--- 368
YY 307 ANSSDMVGRNLMDFGTGSFVASEKLPWGRGPQEMTSLIGFRDGFPRATEAAKKHLSN 366
YY 369 -----RDAFSYGAV--GPEVDSRV-----IVDLRFEGATDPEANNLLVFQNDVODG 412
YY 367 LSRIDQETQKIFKAGLKKPDELQAIQDRSARYVQFCFHEILFQENRIVPSKTATDA 426
YY 413 YSMPQPTFRYPSTASNVRRKMMADMCEV-----ASNILGG-----YLTSPSPQFMDP 460
YY 427 IGI PRPEITY-----AIDYVKGAAHTREVATAKVLGGTVDFNDFANN----- 475
YY 461 GLAHLAGTTRIGFDKATVADNLSNVDFANLYVAGNGTIRTOFGENTPLTSMCHAIS 520
YY 476 ---HITGSTINGADARDSVVDKCRTEHDENLFISSATMPTGTGVNVVTLITIALALRM 531
YY 521 ARSJINTLK 529
YY 532 S-----DTLK 536

RESULT 14
AM48440
D AM48440 standard; Protein; 1209 AA.

XX AC AAM48440;
XX DT 16-JUL-2002 (first entry)
XX DE 2-keto-D-gluconate dehydrogenase.
XX KW Cell membrane bound; 2-keto-D-gluconate dehydrogenase; enzyme;
XX KM 2,5-diketo-D-gluconate.
XX OS Unidentified.
XX FH Key Location/Qualifiers
XX FT Misc-difference 740 /note= "Encoded by TG"
XX PN KR2000019366-A.
XX PD 06-APR-2000.
XX PF 10-SEP-1998; 98KR-0037413.
XX PR 10-SEP-1998; 98KR-0037413.
XX PA (KOAD) KOREA ADV INST SCI & TECHNOLOGY.
XX PI Shin YC, Bahn JG, Yeom DY;
XX DR WPI; 2001-088033/10.
XX DR N-PSDB; ABL52917.
XX PT Base sequence of novel cell membrane-bound 2-keto-D-gluconate
XX dehydrogenase gene and method for conversion into
XX 2,5-diketo-D-gluconate - Noabstract
XX PS Claim 3; Page 9-11; 16pp; Korean.
XX CC The present invention relates to a novel cell membrane-bound
XX 2-keto-D-gluconate dehydrogenase and a method for conversion into
XX 2,5-diketo-D-gluconate. The present sequence is the protein sequence for
XX the 2-keto-D-gluconate dehydrogenase.
XX SQ Sequence 1209 AA;
Query Match 5.9%; Score 170; DB 22; Length 1209;
Best Local Similarity 20.3%; Pred. No. 3.6e-06;
Matches 109; Conservative 65; Mismatches 209; Indels 154; Gaps 21;
QY 68 NGKNPHOREFENLSAEAVTRGVGMSTHTCTSTPRIHP---MESLPGIGRKLSDPAE 124
Db 281 NVTGNADSFQ---QGYLRTVGGTTHWAASCHRRHPSDFVMQSKYGVGR----- 327
QY 125 DDKW---NEL-----YSEAEELIGTSTKEFDESIRHTLVLSLQDAYKDRQIRIFRPLPLA 177
Db 328 ---DWPIQYDDELEPWYCKAENEIGVAGP--NDPARQSPTERS-----QPYPM- 369
QY 178 CHRLKNAPEYVHSAENLPHSIYNDKQKLL-----FTLLTNHRCRTRIAL 223
Db 370 -----DMVPFAGDNYFASVYVPHGYNLVPIQGRSTRPWEGRPTCCGNNNCQICP 421
QY 224 TGGYKIKIGAAEVRNLLATRNPPSOLDSYIM-----AK 256
Db 422 IGAMYNIGIHVE-----PAENGAVVLAEEAVYKMDTSDNNRITAVHMLDTSYGASHKATAK 477
QY 257 VYVLASGAIGNPQILYNSGFGSLQVTPRNDSLIPN-----LGRYITEQPMACQIVLRQEF 312
Db 478 AFALACNGIETPRLL-----NAANDANFNIGIANASDMVGRNMDHSGHCSFLTKE--- 529
QY 313 VDSVRDDPYGLPWKEAVAOHIAKNPT--DALPIPRDPEQVTPPTTEHPHMTQI----- 367
Db 530 -----PVM-----LQKGAQSSCMWGYRDCD-----PRDYSANKVILNNI 565
QY 368 -----HRDAFSYGAVGPEVDSRV-----IVDLRFEGATDPEANNLLVFQNDVODGYSM 415

```

Db      566 SRVTATQQAAMKGLVGVKALDEEIRYRAVHSDLSISLEPLDPDENRLTSLKTRKDPHGL 625
Qy      416 POPTFRYRPSTASNVRAKRWADCEVASNLGGYLPTSPQFMDRGLAL--HLAGTTTRIG 473
Db      626 PCPIIYDVGVYVRKGAASHAQLAHEI-----GQLFDAKEFTISQGLNANNHNGGVIMG 680
Qy      474 FDKATTVAADNNSLVDFANLYVAGNGTIRGTGFGNPTLTSMCHAIKARSIIINTLKG 530
Db      681 KNAKEAVVDGNCRAFDHENLWLPGGGAIPASVNVSTLTMAALGLKAAHDISLRMKG 737

```

RESULT 15

AAB97034
ID AAB97034 standard; Peptide; 34 AA.

XX AC AAB97034;

XX DT 20-JUL-2001 (first entry)

XX DE Lyophyllum shimeji antibacterial protein fragment #4.

XX KW Fungus; antibacterial; antibiotic; plant pathogen; bacterial infection;
XX KW Pyricularia oryzae; Rhizoctonia solani; rice pathogen.

XX OS Lyophyllum shimeji.

XX PN WO200121657-A1.

XX PD 29-MAR-2001.

XX XX 20-SEP-2000; 2000WO-JP06404.

XX XX 21-SEP-1999; 99JP-0267238.

XX XX (NISB) JAPAN TOBACCO INC.

XX XX (NORQ) SOC TECHNO-INNOVATION AGRIC FORESTY & PI.

XX XX Takakura Y, Kuwata S, Inoue Y;

XX DR WPI; 2001-281598/29.

XX PT Antibacterial protein and encoded gene isolated from Lyophyllum

PT shimeji, with activity against plant pathogenic bacteria, applicable in
PT agriculture e.g. rice cultivation at low concentration, produced at low
PT cost on large scale

XX PS Example 2; Page 24; 52pp; Japanese.

XX CC The present sequence is part of an antibacterial protein from the fungus
CC Lyophyllum shimeji. The protein was obtained from a fraction prepared
CC by extracting Lyophyllum shimeji with water and subjecting the extract
CC to ammonium sulphate precipitation. The protein inhibits the growth
CC of the plant pathogenic bacteria Pyricularia oryzae and Rhizoctonia
CC solani at a relatively low concentration. P. oryzae and R. solani are
CC causative of the two major diseases of rice. The protein contains
CC components of 70 kDa and 65 kDa, as determined by SDS-PAGE.

XX CC The antibacterial protein can be produced at low cost on a large scale.

XX SQ Sequence 34 AA;

Query Match 5.8%; Score 168; DB 22; Length 34;

Best Local Similarity 100.0%; Pred. No. 1.4e-08;

Matches 34; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 136 AERLIGTSTKEFDESIRHTLVLSRLQDAYKDRQR 169

Db 1 AERLIGTSTKEFDESIRHTLVLSRLQDAYKDRQR 34

Search completed: October 28, 2003, 15:57:23

Job time : 89 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2003, 15:55:56 ; Search time 29 Seconds
(without alignments)
792.234 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Perfect score: 2901
Sequence: 1 NABEGTAVPVVPGYHKKNEI.....IINTLKGGTDGKNTGHRNL 543

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/2/iaa/5A_COMB.pep.*
- 2: /cgn2_6/prodata/2/iaa/5B_COMB.pep.*
- 3: /cgn2_6/prodata/2/iaa/6A_COMB.pep.*
- 4: /cgn2_6/prodata/2/iaa/6B_COMB.pep.*
- 5: /cgn2_6/prodata/2/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/prodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Match	Length	ID	Description
1	1142.5	39.4	623	1	US-08-734-925-2
2	1119	38.6	622	3	US-09-305-381-2
3	1024.5	35.3	566	3	US-09-023-731-1
4	192	6.6	51	3	US-09-023-731-12
5	159	5.5	615	4	US-09-297-937C-9
6	159	5.5	1276	4	US-09-297-937C-13
7	130.5	4.5	58	3	US-09-023-731-11
8	115	4.0	50	3	US-09-023-731-14
9	108	3.7	35	3	US-09-023-731-15
10	105.5	3.7	778	4	US-09-198-452A-508
11	104.5	3.6	1876	2	US-08-609-049A-12
12	104.5	3.6	1876	3	US-09-170-996-12
13	102.5	3.5	61	3	US-09-023-731-6
14	102.5	3.5	1686	4	US-09-355-160D-2
15	99.5	3.4	618	4	US-09-134-001C-3169
16	96.5	3.3	1876	2	US-08-609-049A-28
17	96.5	3.3	1876	3	US-09-170-996-28
18	95.5	3.3	1257	1	US-08-049-783-2
19	95.5	3.3	1257	1	US-08-158-232-6
20	95.5	3.3	1257	1	US-08-304-626-6
21	95.5	3.3	1257	1	US-08-318-301A-6
22	95.5	3.3	1257	2	US-08-611-928-6
23	95.5	3.3	1257	3	US-09-173-891-6
24	95.5	3.3	1257	3	US-09-076-137-6
25	95.5	3.3	1257	5	PCT-US92-0362A-6
26	95	3.3	3170	2	US-07-642-734C-5
27	95	3.3	3170	3	US-08-439-009A-5

28	92	3.2	585	4	US-09-352-991A-18215	Sequence 18215, A
29	92	3.2	2802	3	US-09-542-331-1	Sequence 1, Appli
30	92	3.2	2802	4	US-09-510-791-1	Sequence 1, Appli
31	91	3.1	890	4	US-09-513-783A-174	Sequence 174, App
32	90.5	3.1	806	3	US-08-549-515-5	Sequence 5, Appli
33	90.5	3.1	806	3	US-08-549-515-11	Sequence 11, Appl
34	90.5	3.1	1313	4	US-09-071-035-450	Sequence 450, App
35	90.5	3.1	1313	4	US-09-071-035-454	Sequence 454, App
36	90	3.1	487	4	US-09-198-452A-907	Sequence 907, App
37	90	3.1	543	3	US-09-199-229-2	Sequence 2, Appli
38	90	3.1	543	3	US-09-443-087-2	Sequence 2, Appli
39	90	3.1	543	4	US-09-687-298-2	Sequence 2, Appli
40	90	3.1	807	3	US-08-974-549A-5	Sequence 5, Appli
41	90	3.1	807	3	US-08-912-951-5	Sequence 5, Appli
42	90	3.1	951	4	US-09-125-635-8	Sequence 8, Appli
43	90	3.1	1132	3	US-08-851-843A-225	Sequence 225, App
44	90	3.1	1132	3	US-08-974-549A-2	Sequence 2, Appli
45	90	3.1	1132	3	US-08-974-549A-344	Sequence 344, App

ALIGNMENTS

RESULT 1
US-08-734-925-2
; Sequence 2, Application US/08734925
; Patent No. 5712139
; GENERAL INFORMATION:
; APPLICANT: NISHIMURA, IKUKO
; APPLICANT: OKADA, KIMIHARU
; APPLICANT: MINAMITHARA, TOMOYUKI
; APPLICANT: KAWAI, GENSUHIRO
; APPLICANT: KOYAMA, YASUJI
; APPLICANT: SUZUKI, MASARU
; TITLE OF INVENTION: NOVEL PYRANOSE OXIDASE, PYRANOSE OXIDASE
; TITLE OF INVENTION: GENE, NOVEL RECOMBINANT DNA AND PROCESS FOR PRODUCING
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
; ADDRESSER: P.C.
; STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, SUITE 400
; CITY: ARLINGTON
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/734,925
; FILING DATE: 22-OCT-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/568,428
; FILING DATE: 06-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: OBLON, NORMAN F.
; REGISTRATION NUMBER: 24,618
; REFERENCE/DOCKET NUMBER: 7126-002-0
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-413-3000
; TELEFAX: 703-413-2220
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 623 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-734-925-2

Query Match 39.4%; Score 1142.5; DB 1; Length 623;
 Best Local Similarity 45.1%; Pred. No. 6.8e-112;
 Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;

QY 13 GYHKKNEIEFKDIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSPGSSAISNGKQP 72
 DB 87 GSHKNTVEYQKNDKFNVIQOLMPVSVPTMTVVDTLSPASQA--STFFVRNGNP 144
 QY 73 HQREFENLSAEAVTRGVGSMTHWTCSTPRIHPHMPESLPGIGRPKL-SNDPAEDDKWNE 131
 DB 145 EQDPLRLNSGQAVTRVVGSMTHWTCATPRPEK-----LQRLLVKNDKADDAWDR 197
 QY 132 LYSEAEERLIGTSTKFEDESIRHTLVLRSLQDAYKDRQIRFPLPLACHRLKNAPPEYVEMH 191
 DB 145 EQDPLRLNSGQAVTRVVGSMTHWTCATPRPEK-----RPLLVKDDQADDAWDR 197
 QY 132 LYSEAEERLIGTSTKFEDESIRHTLVLRSLQDAYKDRQIRFPLPLACHRLKNAPPEYVEMH 191
 DB 198 LYTKAESYFKTGTQCFKESIRHNLVNLKLAEBYKG-QRDFOQIPLAATR--RSPTFFVWS 254
 QY 192 SAENLF--HSIYNDKQKLLFTLLTNHRCRLALTTGGYEKKIGAAEVNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPTDAPNERFNLPPAVACRY-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKVYVLAGAIGNPOILYNSGSGI-QVTPRN-DSLIPNLGRY 295
 DB 296 ESHLHIDLIISGRFEIKADVFVLTAGAVHNAQLLVNSGFGOLGRPDPPANPQLPSLSGY 355
 QY 296 ITEQPMAPCOIVLROEFVDSVRDDPY--GLP-----WKKEAVAQH 333
 DB 356 ITEQSLVFCQVMTSLDVSXSDMIIRGNFGDLGYSVITYPGAETKNKHPDWNKVKQH 415
 QY 334 IAKNFTDALPIPRDPPEQVTTPTTEHPHMTQIHRDAFSYAGVGVDSRVIVDLRWF 393
 DB 416 NMHQEDPLPIPFEPPEQVTTLPQSHPMHTQIHRDAFSYAGVGVDSRVIVDLRWF 475
 QY 394 ATDEANLLVPONDVQSGNSPQPTFRYR--PSTASNVARKMADMCVANSNGLGYLPT 452
 DB 476 RTEPKEENKLFSDKITDTYNNPQPTDFRFPAGRTSKEAEDMTDMCVMSAKIGGFLPG 535
 QY 453 SPQPMDFGLALHLAGTTRIGFDKA--TTVADNNSLVWDFANLYVAGNGTIRTGFGENTP 510
 DB 536 SLPQMEFGLVHLGGTHRMGFDEQEDKCCVNTDSRVFGKNLFLGGCGNIPYAGNPT 595
 QY 511 LTSMCHAISARSIN 526
 DB 596 LTAMSLAISKSEYIKN 611

RESULT 2

US-09-305-381-2
 ; Sequence 2, Application US/09305381
 ; Patent No. 6146865
 ; GENERAL INFORMATION:
 ; APPLICANT: Christensen, Soren
 ; APPLICANT: Lassen, Soren Plensted
 ; APPLICANT: Schneider, Palle
 ; TITLE OF INVENTION: Nucleic Acids Encoding Polypeptides
 ; FILE REFERENCE: 5571-200-US
 ; CURRENT APPLICATION NUMBER: US/09/305,381
 ; EARLIER FILING DATE: 1999-05-05
 ; EARLIER APPLICATION NUMBER: 60/088,724
 ; EARLIER FILING DATE: 1998-06-10
 ; EARLIER APPLICATION NUMBER: PA 1998 00774
 ; NUMBER OF SEQ ID NOS: 2
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 622
 ; TYPE: PRT
 ; ORGANISM: Coriolus Versicolor
 US-09-305-381-2

Query Match 38.6%; Score 1119; DB 3; Length 622;
 Best Local Similarity 44.6%; Pred. No. 2.1e-109;

Matches 251; Conservative 69; Mismatches 183; Indels 60; Gaps 15;

QY 13 GYHKKNEIEFKDIDRFVNVIKGALQOVSVPRNQNVPTLDPGAWSPGSSAISNGKQP 72
 DB 87 GSHKNTVEYQKNDKFNVIQOLMPVSVPTMTVVDTLSPASQA--STFFVRNGNP 144
 QY 73 HQREFENLSAEAVTRGVGSMTHWTCSTPRIHPHMPESLPGIGRPKL-SNDPAEDDKWNE 131
 DB 145 EQDPLRLNSGQAVTRVVGSMTHWTCATPRPEK-----LQRLLVKNDKADDAWDR 197
 QY 132 LYSEAEERLIGTSTKFEDESIRHTLVLRSLQDAYKDRQIRFPLPLACHRLKNAPPEYVEMH 191
 DB 198 LYTKAESYFKTGTQCFKESIRHNLVNLKLAEBYKG-QRDFOQIPLAATR--RSPTFFVWS 254
 QY 192 SAENLF--HSIYNDKQKLLFTLLTNHRCRLALTTGGYEKKIGAAEVNLLATRNPSQL 249
 DB 255 SAHTVFDLQNRPTDAPNERFNLPPAVACRY-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKVYVLAGAIGNPOILYNSGSGI-QVTPRN-DSLIPNLGRYITEQPMAP 303
 DB 311 --IKAVVYILTAGAVHNAQLLAASFGQL--GRPDPAKPLPSLLPYLGLTHITEQTLVF 364
 QY 304 COIVLROEFVDSVRDDPY--GLP-----WKKEAVAQHIKKNPTDA 341
 DB 365 CQVMTSLINSVTADMTIVGKPGHPDYSVITYPGAETKNKHPDWNKVKHMDHQEDP 424
 QY 342 LPIPRDPPEQVTTPTTEHPHMTQIHRDAFSYAGVGVDSRVIVDLRWFATDPEANN 401
 DB 425 LPIPFEPPEQVTTLPQSHPMHTQIHRDAFSYAGVGVDSRVIVDLRWFATDPEANN 484
 QY 402 LLVFQNDVQSGNSPQPTFRYR--PSTASNVARKMADMCVANSNGLGYLPTSPQPMDFG 461
 DB 485 KLFPSDKITDAYNLROQTFDPR--FPGREAEEDMTDMCVMSAKIGGFLPGSYQPMDFG 542
 QY 462 LAHLACTTIRIGFDKA--TTVADNNSLVWDFANLYVAGNGTIRTGFGENTPILTSCHAIK 519
 DB 543 LVLHLGGTHRMGFDEKADKCCVNTDSRVFGKNLFLGGCGNIPYAGNPTLTAMSLA 602
 QY 520 SARSINLTKGGTDGKNTGEHN 542
 DB 603 SCEYI---KKNFEPSPNPVKHN 622

RESULT 3

US-09-023-731-1
 ; Sequence 1, Application US/09023731
 ; Patent No. 6291648
 ; GENERAL INFORMATION:
 ; APPLICANT: Kawamura, Yukio; Morita,
 ; APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
 ; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
 ; TITLE OF INVENTION: GENE ENCODING SAME
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: MICROSOFT WORD 97
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/023,731
 ; FILING DATE: 13-FEB-1998
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: JP 29275/1997
 ; FILING DATE: 13-FEB-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: KENNETH H. SONNENFELD

REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 566
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-023-731-1

Query Match 35.3%; Score 1024.5; DB 3; Length 566;
Best Local Similarity 43.0%; Pred. No. 1.9e-99;
Matches 232; Conservative 77; Mismatches 180; Indels 51; Gaps 12;

2 AEGTAVPVVPGVHKHKNTEIEFKQIDRFVNVKALQVSVPRNQNVPTLDPGAWAPP 61
62 AEGSDNPVIGAHKNSIKFKQIDKFNINIGALQISPSDYQTLVAW-APP 120
62 GSSA-----ISNGKPHQREFENLSAEAVTRGVGMSHTWCSTPRIHPWESLPGIGREK 117
121 IDPAEQQLVIMGNPNQNEAGLNLPGSAVTRTVGGMATHWCACPTPHD-----EE 170
118 LSNDAEDDKENNELYSEARLIGTSTKEFDSIRHTLVLSLQDAYKDRQIRFRLPLA 177
171 RVNPNV-DKQEPDALLERAKTLNHNHSDQYDSDIRQIVVKETLQOTL-DASRGVTTPLG 228
178 CHRLKNAPEYVHWAENLPHSIYNDKQKLFLLTNHRCRLALTGGYEKKIGAEVR 237
229 VERRDNIYVWTGAD-----TVLGDVPSRPFALVTETRTVKLVSETNPTQVVAALLR 284
238 NLLATRNPSQSDYSIMARKVYVYLASGAIGNPOLXNSGFSGLQVTPRNDLIPNLGRYIT 297
285 NL-----NTSN--DELVAKSFVACGAVCTPOLNWS-----NIRPY-----ALGRVLS 327
298 EOPMAFCQIVLRQEFVSDVRDDPYGLPWKEAVAQVIAKNTDLPIDPRDPQVTPPF 357
328 EQSMFTFCQIVLRGIVDAIATDPR---FAAKVEAKKHGPDVLPPIRHEPEQWMI 383
358 TEHPHWTQIHRDAFSYAGVGEVDSRVIVDLRWGATDPEANNLLVF-----QND 408
394 TSDPFVHVQVHRDAFSYGVGKADPRVVVDLRFEGKSDIVEENRTVFGPNKLRWEAG 443
409 VQDGSMFQPTFRYRPSSTASVRAKMMADMCEVASNLGGYLPTSPQFMDPLALHLAG 468
444 VTDYGMFQPTFRVKTWADGDRDQRMNDMTNVANMLGGYLPQFQFMAPGLVHLITG 503
469 TTRIGFDKATTVADNNSLVDFANLYVAGNGTIRTGFGENPFLTSMCHAIKSARSINTL 528
504 TTRIGTDQTSVADPTSKVHNFNNLWVGNGCIPDATACNPTRTSVAYALKGAEEVVNL 563

RESULT 4
US-09-023-731-12
Sequence 12, Application US/09023731
Patent No. 6291648
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 51
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-12

Query Match 6.6%; Score 192; DB 3; Length 51;
Best Local Similarity 62.7%; Pred. No. 5.5e-13;
Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;

QY 330 VAOHAKNPTDALPIPRDPPEQVTPTEHPHWTQIHRDAFSYAGVGE 380
DB 1 VEAHKKHPDVLPIRHEPEQWMIPTSDPFVHVQVHRDAFSYGVGPK 51

RESULT 5
US-09-297-937C-9
Sequence 9, Application US/09297937C
Patent No. 6337199
GENERAL INFORMATION:
APPLICANT: YUM, Do Young
APPLICANT: PAN, Jae Gu
TITLE OF INVENTION: Membrane-Bound Glucanase Dehydrogenase, Gene Sequence
TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate
TITLE OF INVENTION: Using Transformed Recombinant E. Coli
FILE REFERENCE: P66159USO
CURRENT APPLICATION NUMBER: US/09/297,937C
CURRENT FILING DATE: 1999-05-11
PRIOR APPLICATION NUMBER: PCT/KR98/00296
PRIOR FILING DATE: 1998-09-25
PRIOR APPLICATION NUMBER: KR 97-48802
PRIOR FILING DATE: 1997-09-25
NUMBER OF SEQ ID NOS: 13
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 615
TYPE: PRT
ORGANISM: Erwinia cyripedii
US-09-297-937C-9

Query Match 5.5%; Score 159; DB 4; Length 615;
Best Local Similarity 19.9%; Pred. No. 1.3e-07;
Matches 124; Conservative 85; Mismatches 257; Indels 158; Gaps 32;

QY 4 EGTAVPVVPGVHKHKNTEIEFKQIDRFVNV-----KCALQVSVPRNQNVPTLDPGAW 58
DB 5 ERVSVV-VSGVSRGEGVTVANLKKVDVAVVGVFGAGAINAKELTEAGLNVALERG--- 60
QY 59 APGSSAISNGKNP-----HOREFENLSAEVT-----R 87
DB 61 --PHRTYPGAYPQSIDELTYNIRKKLFQLSKSTVIRHDASQTAVPYRQLAFLPGT 118
QY 88 GVGMSTHTWCTSTPRIHPMESLPG-----IGRPKSLNDPAEDD--KEMNEL---YSEAE 137

```

119 GTGAGLHWSGVHFRVDPVELNLSHVEARYGKFIPEGMTIQDFGVSYNELEFPFDOAE 178
138 RLIGTSKEDFDESIRHTLVRLSLQDAY--KDRQIRFRPLP-----LACHRLKNAPEYVEW 190
179 KVFCTSGSAM--TIKGMIGKEKGGNFYAPDRSSDF--PLPAQKRITYSAQLFAQAASVGY 235
191 H-----SAENLFHSIYND-----KOKKLF 212
236 HPYDMP SANTSGPYNTYGAQMGPCNFCGVCYACVMYSKASPNVNLPALEQEPKPEL 295
213 LTNHRCRLALTGGEYKKGICA---AEVRNLLATRNPSO--LDSYIMAKVYVYLASGAIG 266
296 RNNAYVLRNLTGDKKATGVTYLDGQREV---QPADLVLSAFOFHNHMLLSGIG 352
267 NP--QILYNSGFGSLQVTPRNDLIPNL--GRYTEQPM---AFQOIVLRQBFVSDVRDDPY 331
353 QPVNPIITNEGTVGRNFAYQNI STLKALFDKNTTTPF IGAGGAGVAVDDFNADNPDHGPY 412
322 GL-----PMWKEAVAQHIKNTDALTPIPRDPEQVTTPTTEHPHWHQIHRDAFSYGAV 377
413 GFVGSGSPFW---VNOAGTKPVSGLEPTKGTN-----WGSQWK-----AAV 450
378 GPEVDSRVIVDLRWFGATDPEANLLVFQNDVODGYSM--OPTFRYRPSSTASNVR--ARK 434
451 ADTYNHHSIDAH--GAHOSYRANYLDLPYKNVYGOPLLRMTFDWQD---NDIRMAQF 505
435 MMADMCEV--ASN-----LGGYLPSTPPQFMDPGL--ALHLAGTRIGFDKATTVADNNSL 486
506 MVGKMKRKITAMNPKMIIGG--AKGPGTHFTDITVYQTHMSGGAINGEDPKTSAVNRYLQ 563
487 VMDPANLYVAGNGTIRTGFCENPT 510
564 SWDVPNVFVPGASAPFQGLGYNPT 587

```

RESULT 6

```

US-09-297-937C-13
; Sequence 13, Application US/09297937C
; Patent No. 6337199
; GENERAL INFORMATION:
; APPLICANT: YUM, Do Young
; APPLICANT: PAN, Jae Gu
; TITLE OF INVENTION: Membrane-Bound Gluconate Dehydrogenase, Gene Sequence
; TITLE OF INVENTION: Encoding the Same and Production of 2-Keto-D-Gluconate
; TITLE OF INVENTION: Using Transformed Recombinant E. Coli
; FILE REFERENCE: P66159USO
; CURRENT APPLICATION NUMBER: US/09/297,937C
; PRIOR FILING DATE: 1999-05-11
; PRIOR APPLICATION NUMBER: PCT/KR98/00296
; PRIOR FILING DATE: 1998-09-25
; PRIOR APPLICATION NUMBER: KR 97-48802
; PRIOR FILING DATE: 1997-09-25
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 1276
; TYPE: PRT
; ORGANISM: Erwinia cyripedii
US-09-297-937C-13

```

```

Query Match 5.58; Score 159; DB 4; Length 1276;
Best Local Similarity 19.9%; Pred. No. 4.6e-07;
Matches 124; Conservative 85; Mismatches 257; Indels 158; Gaps 32;

Qy 4 EGTAVPVVPGYHKKNEIEFOKIDRFVNVJ-----KGAQQSVVPVNRQNVPTLDPGAW 58
Db 225 ERVSVP-VSGYSRGEVTVANELKKYDVAVVVGWAGWAKELTGLNVVALERG---280
Qy 59 APGSSAISNGKVP-----HQREFNLGAEAVT-----R 87
Db 281 --PHRDTPDGAVPQSIDELTYNIRKLFQDLKSTVTIRHDASQTAVPYRQLAAFLPGT 338

```

```

Qy 88 GVGGMSTHWTCTPRHPWESLPG-----IGRPKLSNDPAEDD--KENNEL---YSEAE 137
Db 339 GTGGAGLHWSGVHFRVDPVELNLSHVEARYGKFIPEGMTIQDFGVSYNELEFPFDOAE 398
Qy 138 RLIGTSKEDFDESIRHTLVRLSLQDAY--KDRQIRFRPLP-----LACHRLKNAPEYVEW 190
Db 399 KVFCTSGSAM--TIKGMIGKEKGGNFYAPDRSSDF--PLPAQKRITYSAQLFAQAASVGY 455
Qy 191 H-----SAENLFHSIYND-----KOKKLF 212
Db 456 HPYDMP SANTSGPYNTYGAQMGPCNFCGVCYACVMYSKASPNVNLPALEQEPKPEL 515
Qy 213 LTNHRCRLALTGGEYKKGICA---AEVRNLLATRNPSO--LDSYIMAKVYVYLASGAIG 266
Db 516 RNNAYVLRNLTGDKKATGVTYLDGQREV---QPADLVLSAFOFHNHMLLSGIG 572
Qy 267 NP--QILYNSGFGSLQVTPRNDLIPNL--GRYTEQPM---AFQOIVLRQBFVSDVRDDPY 321
Db 573 QPVNPIITNEGTVGRNFAYQNI STLKALFDKNTTTPF IGAGGAGVAVDDFNADNPDHGPY 632
Qy 322 GL-----PMWKEAVAQHIKNTDALTPIPRDPEQVTTPTTEHPHWHQIHRDAFSYGAV 377
Db 633 GFVGSGSPFW---VNOAGTKPVSGLEPTKGTN-----WGSQWK-----AAV 670
Qy 378 GPEVDSRVIVDLRWFGATDPEANLLVFQNDVODGYSM--OPTFRYRPSSTASNVR--ARK 434
Db 671 ADTYNHHSIDAH--GAHOSYRANYLDLPYKNVYGOPLLRMTFDWQD---NDIRMAQF 725
Qy 435 MMADMCEV--ASN-----LGGYLPSTPPQFMDPGL--ALHLAGTRIGFDKATTVADNNSL 486
Db 726 MVGKMKRKITAMNPKMIIGG--AKGPGTHFTDITVYQTHMSGGAINGEDPKTSAVNRYLQ 783
Qy 487 VMDPANLYVAGNGTIRTGFCENPT 510
Db 784 SWDVPNVFVPGASAPFQGLGYNPT 807

```

RESULT 7

```

US-09-023-731-11
; Sequence 11, Application US/09023731
; Patent No. 6291648
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; APPLICANT: Akihiko; Izumo, Koji; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; TITLE OF INVENTION: GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/023,731
; FILING DATE: 13-FEB-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792

```

```

RESULT 10
US-09-198-452A-508
; Sequence 508, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments thereof and uses thereof, in particular for the diagnosis, prevention and treatment of infection
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
; SEQ ID NO 508
; LENGTH: 778
; TYPE: PRT
; ORGANISM: Chlamydia pneumoniae
US-09-198-452A-508

```



```
Query Match      3.7%; Score 106.5; DB 4; Length 778;
Best Local Similarity 20.9%; Pred. No. 0.073;
Matches 84; Conservative 59; Mismatches 158; Indels 101; Gaps 20;

QY 116 PKLNDPAEDDKENWELYSERLIGTSTKEFDE-----SIRHTLVLSLQDA-----Y 164
DB 446 PRRNGSPREDSPLMNALVGNHAK-HGAKTKESSSTPEISISAPIVRGWSQDSVSFIV 504
QY 165 KDRQRIFRPLPLACHRLKNAPEYVWHSANLPHSIYNDKQKFLTLTNHRCRLALT 224
DB 505 MEDDHIFVDVPRKXGIDYDVPSPRWSGPARELEEDVEFD----- 543
QY 225 GYKKGKGAABVRNLLATRNPSQLDSYIMAKVYVLASGAIGNPQIILYNSGFSGLQVTPR 284
DB 544 --YEVPTSAB-----PSKDKNIYMTPR---LATPAIYD--LPSRPGSGSSRSRPS 587
QY 285 ND---SLIPNLGRVITEOPMAFCQIVLRQEPVDSVRDPPVGLPWKEAVACHIAKNPTDA 341
DB 588 SDRVRSSPN--RRGVPLPPVP--SPAMSEE--GSIYEDMSGAGAGESDYEDMSRSPS-- 640
QY 342 LPIDFRDPEQVTTPTTEHPWHTQIHRDAF-----SYGAVGPEYDSRVIVDLRMFGATDP 397
DB 641 ---PRGDLDEPIYANTPEDNPF--TORNIDRLQERSGASASVPEP-IYDEIPWIGHRPP 695
QY 398 EANNLLVFQNDVQGYCMPQP-----TFRYRSTASNVRRARQOMADMCEVASNLGGYL 450
DB 696 A-----TLPRPENTLTNVLRSVSGFGPEVRAALLSESVGAVMVEARSIV 740
QY 451 PTSPPP---QFMDDP--GLALHLAGTTRIGDKATTVDNNS 485
DB 741 PPTFPGGSESYLEPLGG-----LVATTKILLOKGNPRGESNA 778

RESULT 11
US-08-609-049A-12
; Sequence 12, Application US/08609049A
; Patent No. 5948664
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 5948664el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/609,049A
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1876 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein

Query Match      3.6%; Score 104.5; DB 2; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.55;
Matches 85; Conservative 53; Mismatches 170; Indels 79; Gaps 21;

QY 8 VPVPGYHKKNVETEFQKIDRFVNVIKGALQOVSVPVRNQNVPTLDRGA-----WSAP--P 61
DB 184 VVQPAAQOORPLN--SEELQRLYSM--PAQMAVVPVQPNAYMYFGAVVTPYTAIPLV 239
QY 62 GSAISNGKNPHOREPENLSAEAVTRGVGGMSTHTCTSPRIHP--PMESLPGIG--RPKL 118
DB 240 GSAAPWPPQYPAQ-----GYGFGAYTHMDLRRPQSQAPAPQQTATTSHHSQP 288
QY 119 SNDAEDDKENWELYSER---LIGTSTKEFDESIRHTLVLSLQDA-----YKDRQRI 170
DB 289 SNHSTSPAEANGVAFPARQVPSTGVSSSSHTGNGHSSVPRRGNLDLIDNHEDYSRV 348
QY 171 -----FRPLPLACHRLKNAPEYVWHSANLPHSIYNDKQKFLTLTNHRCRLALTG 225
DB 349 SVLEAFDPLNDNTGNDTASDSTSYAAYDPDFLYSGD-----AATQYSDPMYEAVN 401
QY 226 GYKKGKGAABVR--NLLATR-----NPSSQLDSYIMA-----KYVVLASGAIGNP-- 268
DB 402 RWDKTATVSPNVGLIGRWODFLSQSPSTSSSQYGVAPPEESLKLAEANGSGTISPPPLPP 461
QY 269 --OILYNSGFSGLQVT--PRNDSLIPNLGRVITEOPMAFCQIVLRQ-----EFVDSVR 317
DB 462 RNQOCYESNOAMPVSRPQSSVLTD--SYTSSIP--ANVLDRRRTCTRLYELISDQR 516
QY 318 -DDPYGLPWKEAVAQHIANKNPTDALP 343
DB 517 TDDPELLEFY-HMVKEVRARYPHDDAP 542

RESULT 12
US-09-170-996-12
; Sequence 12, Application US/09170996
; Patent No. 6291220
; GENERAL INFORMATION:
; APPLICANT: Williams, Lewis T.
; APPLICANT: Molz, Lisa
; APPLICANT: Chen, Yen-Wen
; TITLE OF INVENTION: No. 6291220el PI 3-Kinase Polypeptides
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/170,996
; FILING DATE: 29-FEB-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Karen B.
; REGISTRATION NUMBER: 29,684
; REFERENCE/DOCKET NUMBER: 2307K-063700US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-326-2400
; TELEFAX: 415-326-2422
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
```

LENGTH: 1876 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-170-996-12

Query Match 3.6%; Score 104.5; DB 3; Length 1876;
Best Local Similarity 22.0%; Pred. No. 0.55; 170; Indels 79; Gaps 21;
Matches 85; Conservative

QY 8 VPVVGHHKNEIEFQKIDRFNVKIGALQQVSVPRNQNVPTLDPCA-----WSAP--P 61
DB 184 VPQPAQAQQORPLN-SEELQRLYSM---PAQMAVVPVQPNAYMYGAVVPTATPIVP 239
QY 62 GSSAISNGKNPHQREFENLSAEAVTRGVGGMSTHTWCSTPRIHP-PMESLPGIG--RPKL 118
DB 240 GSAAFMPPOQPAQ-----GYGFGGAYTHMDLRRPQSQPAPQQTAPTTHSHHSQP 288
QY 119 SNDAEDDKENNELYSAEER-----LICTSKPEDESIRHTLVLSLODA-----YKDRQRI 170
DB 289 SNHSTSSPAEANGVAPPARQVSTGVSSSSHTGNGHSSVPRRGNLDLIDLNHEDYSRV 348
QY 171 -----FRPLPLACHRLKNAPEYVEHSAENLFHSIYNDKQKGLFTLLTNHRCRLALTG 225
DB 349 SVLEAFDPLNDNTGNTASDTSYAEYDPDFLYSGD-----AATQSDPMVEAVN 401
QY 226 GYEKKGIAEVR-NLIATR-----NPSQLDSYIMA-----KYVVLASGAIGNP----- 268
DB 402 RWDKTATVSPNVGLIGWRQDFLSQSFSSQYGVAPPEBSLKLAEANGSOTISPPPLPP 461
QY 269 -OILVNSGFGSLQVT-PRNDSLIPNLGRYITROPMAFCQIVLRO-----EFVDSVR 317
DB 462 RNCQCVESNOAMPVSRPOSSVLTD--SYTSSIP---ANVLDRRKTCRTRYELISDQR 516
QY 318 -DDPYGLPMWKEAVAQIAKNPDTDALP 343
DB 517 TDDPELLEFY-HMWKEVRARYPHDDAP 542

RESULT 13
US-09-023-731-6
Sequence 6, Application US/09023731
Patent No. 6251648
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita,
APPLICANT: Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTI-TUMOR PROTEIN AND
TITLE OF INVENTION: GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/023,731
FILING DATE: 13-FEB-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 61
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-023-731-6

Query Match 3.5%; Score 102.5; DB 3; Length 61;
Best Local Similarity 43.3%; Pred. No. 0.0023;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;

QY 29 FVNVIKGAQQVSVPRNQNVPTLDPCAWSAPPGSSA-----ISNGKNPHQREFENLSAE 84
DB 1 FVNIINGALQPISSISPSDYQFTPLAAM-APPIDPAEQOLVINGHNPQOEAGLNLP 59
QY 85 VT 86
DB 60 VT 61

RESULT 14
US-09-355-160D-2
Sequence 2, Application US/09355160D
Patent No. 6436671
GENERAL INFORMATION:
APPLICANT: Domin, Jan
TITLE OF INVENTION: No. 6436671el Lipid Kinase
FILE REFERENCE: 2332-1-004
CURRENT APPLICATION NUMBER: US/09/355,160D
CURRENT FILING DATE: 1999-10-01
PRIOR APPLICATION NUMBER: PCT/GB98/00244
PRIOR FILING DATE: 1998-01-27
PRIOR APPLICATION NUMBER: 9701652.1
PRIOR FILING DATE: 1997-01-28
NUMBER OF SEQ ID NOS: 11
SOFTWARE: Patent in version 3.1
SEQ ID NO 2
LENGTH: 1686
TYPE: PRT
ORGANISM: Homo sapiens
US-09-355-160D-2

Query Match 3.5%; Score 102.5; DB 4; Length 1686;
Best Local Similarity 24.3%; Pred. No. 0.74;
Matches 74; Conservative 41; Mismatches 116; Indels 73; Gaps 19;

QY 50 PTLDPGAMSAPPG-----SSAISNGKNPHQREFENLSAEAVTRGVG 90
DB 135 PTIQRGQW--PPGLPGPSTYALPSIYPTYSKQAAQFQNGFPRMPTFP--STEPYLSLP 190
QY 91 GMSTHW-----TCSTPRIHPMESLPGIGRKLNSNDPAEDDKENNELYSAEERL----ICT 142
DB 191 GSPYFSYPLTPATP-PH-PQGSLLP-IYRPVSTDMA---KLFDKIATSEFLKNGKART 244
QY 143 STKEPDESIRHTLVLSLODAYKORQIRFRL-----PLACHRLKNAPEYVEHSAENLFHS 199
DB 245 DLEITDSKVSNLQVSPKSEDISK-----FDWLDLPLSKPKVDNV-EVLDEHEEKVSSL 298
QY 200 IYNDKQKGLFTLLTNHRCRLALTGVEKKGIAEAVENLLATRNPPSSQLDSYIMAKYVV 259
DB 299 LAKDPWDAVLBERSTANC-----HLERKVGKSGSVATVTRSQSLNIRTTOLAK--- 348
QY 260 LASGAIGNPQILYNSGFGSLQVTPRNDSLIPNLGRYITEQPMAFQIV--LRQBFVDSVR 317
DB 349 -AQGHISQKD---PNGTSSL---PTGSSLLQEV-EVQNEENAAFCRSITKLTKEF----- 395
QY 318 DDPY 321
DB 396 --PY 397

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2003, 15:59:57 ; Search time 72 Seconds
(without alignments)

1262.932 Million cell updates/sec

Title: US-09-856-327-2_copy_76_618

Perfect score: 2901

Sequence: 1 NAEAGTAVYVGVYHKNEI.....IINTLKGDTGKNTGHRNL 543

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 629382 seqs, 167460630 residues

Total number of hits satisfying chosen parameters: 629382

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.*

1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2: /cgn2_6/ptodata/2/pubpaa/FCR_NEW_PUB.pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/2/pubpaa/FCRUS_PUBCOMB.pep.*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep.*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep.*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep.*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep.*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	1024.5	35.3	566	10	US-09-911-969-1
2	192	6.6	51	10	US-09-911-969-12
3	130.5	4.5	58	10	US-09-911-969-11
4	129	4.4	545	15	US-10-156-761-9468
5	115	4.0	50	10	US-09-911-969-14
6	108	3.7	35	10	US-09-911-969-15
7	106.5	3.7	775	11	US-09-920-843A-58
8	106	3.7	546	13	US-10-253-971-3
9	104	3.6	514	15	US-10-156-761-9160
10	103.5	3.6	342	15	US-10-156-761-14017
11	102.5	3.5	61	10	US-09-911-969-6
12	102.5	3.5	1686	14	US-10-092-219-2
13	102	3.5	1244	9	US-09-815-915-8
14	102	3.5	1244	12	US-10-393-316-8
15	97.5	3.4	1536	10	US-09-801-368-322

16	97	3.3	804	15	US-10-128-714-3088	Sequence 3088, Ap
17	97	3.3	1028	15	US-10-128-714-8088	Sequence 8088, Ap
18	96	3.3	970	15	US-10-026-021-2	Sequence 2, Appli
19	95.5	3.3	1257	9	US-09-738-363-6	Sequence 6, Appli
20	95	3.3	1778	12	US-10-238-075-749	Sequence 749, App
21	94	3.2	1286	15	US-10-156-761-14323	Sequence 14323, A
22	94	3.2	1714	15	US-10-128-714-3176	Sequence 3176, Ap
23	94	3.2	1750	15	US-10-128-714-8176	Sequence 8176, Ap
24	92	3.2	733	12	US-10-291-583-60	Sequence 60, Appl
25	92	3.2	1381	12	US-10-032-585-7784	Sequence 7784, Ap
26	91.5	3.2	299	10	US-09-764-864-1522	Sequence 1522, Ap
27	91.5	3.2	350	10	US-09-764-864-1102	Sequence 1102, Ap
28	91.5	3.2	757	15	US-10-014-789A-4	Sequence 4, Appli
29	91	3.1	568	10	US-09-873-880-24	Sequence 24, Appl
30	91	3.1	1390	15	US-10-100-957A-174	Sequence 174, App
31	91	3.1	890	15	US-10-097-340-35	Sequence 35, Appl
32	90.5	3.1	847	15	US-10-177-293-498	Sequence 498, App
33	90	3.1	492	10	US-09-881-752A-142	Sequence 142, App
34	90	3.1	807	15	US-10-044-692-5	Sequence 5, Appli
35	90	3.1	951	12	US-10-044-539-5	Sequence 5, Appli
36	90	3.1	1132	10	US-09-990-080-2	Sequence 2, Appli
37	90	3.1	1132	10	US-09-749-728B-31	Sequence 31, Appl
38	90	3.1	1132	10	US-09-843-676-225	Sequence 225, App
39	90	3.1	1132	10	US-09-953-052-2	Sequence 2, Appli
40	90	3.1	1132	12	US-10-295-681-57	Sequence 57, Appl
41	90	3.1	1132	15	US-10-053-758-225	Sequence 225, App
42	90	3.1	1132	15	US-10-208-243-2	Sequence 2, Appli
43	90	3.1	1132	15	US-10-054-295-225	Sequence 225, App
44	90	3.1	1132	15	US-10-054-611-225	Sequence 225, App
45	90	3.1	1132	15	US-10-054-611-225	Sequence 225, App

ALIGNMENTS

RESULT 1

US-09-911-969-1
; Sequence 1, Application US/09911969
; Patent No. US20020137896A1
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; Akihiro; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: MICROSOFT WORD 97
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/911.969
; FILING DATE: 24-Jul-2001
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/023.731
; FILING DATE: 13-FEB-1998
; APPLICATION NUMBER: JP 29275/1997
; FILING DATE: 13-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: KENNETH H. SONNENFELD
; REGISTRATION NUMBER: 33,285
; REFERENCE/DOCKET NUMBER: 3479-4000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 758-4800
; TELEFAX: (212) 751-6849
; TELEX: 421792

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 566
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-911-969-1

Query Match 35.3%; Score 1024.5; DB 10; Length 566;
Best Local Similarity 43.0%; Pred. No. 3.4e-90;
Matches 232; Conservative 77; Mismatches 180; Indels 51; Gaps 12;
QY 2 AEEGTAVYVPGYHKKNIEFQKIDIRFNNIKALQVSVPRNQNVPTLDPCAWAPP 61
DB 62 AEIGSQDNPIVGAHKNSIKFKQIDKFNINGALQISPSDSTQYPTLVAAM-APP 120
QY 62 GSSA---ISGNKPHOREFENLSAEAVTRGVGGMSTHTWCTSTRIHPMESLFGIGRPK 117
DB 121 IDPAEGQLVINGHNPQAGLNPQSAVTRVGGNATHWTCACTPHD-----EE 170
QY 118 LSNDAEDDKENNELYSABERLIGTSTKEFDESIRHTLVLSLQDAYKDRQRIERPLPLA 177
DB 171 RVNPNV-DKQEFALLERAKTLNVHSDQYDDSIQIVVKETLQOTL-DASRGVTTILPG 228
QY 178 CHRUKNAPYEVWESAENLFHSIYNDKQKFLTLTNHRCRTRALTGCGYKKIGAAEVR 237
DB 229 VERRTDNPIYVTTGAD---TVLGDVPKSPRFALVTETRTVLIVSETNPTQVVAALLR 284
QY 238 NLLATRNPSQDLSYMAKVVLVLAGAIGNQILVNSGFGSLQVTPRNDLSLIPNLGRYIT 297
DB 295 NL-----NTSN--DELVAKSVIAGAVCTQQLWNS-----NIRPY-----ALGRYLS 327
QY 298 EQPMAFQVILVRQFVDSVRDDPYGLPMWKEAVAQIAKNTDLPPIFRDPEPQVTTTF 357
DB 328 EQSMTFCQVILKRGIVDAIATDPR----FAAKEVAHKKKHDDVLPPIPFHEPEQVMPY 383
QY 358 TEEHPWHTQIHRDAFSGAVGPEVDSRVIVDLRFMGATDPEANLLVF-----QND 408
DB 384 TSDPFMHVQVHRDAFSGVGVGPKADPRVVVDLRFEGKSDIVEENRVTFGPNPKLREWEAG 443
QY 409 VODGVSMPQPTFRYRPSFASVNRARKMADNCEVASNLGGLYPTSPQFMDPGLALHLAG 468
DB 444 VTDYGMQPTFHVKRTNADGDRQNMNDTNVNLGGYLPQSGYPOFVAPGLVHLHTG 503
QY 469 TTRIGFDKATTVADNNSLVMPFAMLYVAGNCTIRTFGENPTLTSMCHAIKARSASINTL 528
DB 504 TTRIGTDDTSVADPTSKVHNFNLVWGGNCIPDATACNPTRTSVAYALKGAEVNVYL 563

RESULT 2
US-09-911-969-12
Sequence 12, Application US/09911969
Patent No. US20020137896A1
GENERAL INFORMATION:

APPLICANT: Kawamura, Yukio; Morita,
Akhihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
GENE ENCODING SAME

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/911,969
FILING DATE: 24-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 51
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-911-969-12

Query Match 6.6%; Score 192; DB 10; Length 51;
Best Local Similarity 62.7%; Pred. No. 4e-11;
Matches 32; Conservative 6; Mismatches 13; Indels 0; Gaps 0;
QY 330 VAQHIANKPTDALPIFRDPEQVTTPTTEHPWHTQIHRDAFSGYGVGPE 380
DB 1 VEAHKKHDPDVLPIPFHEPEQVMPYTSDFPHVQVHRDAFSGYGVGPK 51

RESULT 3

US-09-911-969-11
Sequence 11, Application US/09911969
Patent No. US20020137896A1
GENERAL INFORMATION:

APPLICANT: Kawamura, Yukio; Morita,
Akhihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND
GENE ENCODING SAME

NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154

COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/911,969
FILING DATE: 24-Jul-2001

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 751-6849
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:

```

; LENGTH: 58
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-911-969-11

Query Match      4.5%; Score 130.5; DB 10; Length 58;
Best Local Similarity 44.4; Pred. No. 4.6e-05;
Matches 28; Conservative 11; Mismatches 13; Indels 11; Gaps 2;

258 YVLASGAGNPGQIIYNVSGFSLQVTPRNDSLINPLGRVITQPMACQIVLRFQFVDSVR 317
ib 2 FVIAAGVCTPIQLWNS-----NIRPY-----ALGRVYSEQSMFPCQIVLKRGVDAIA 50
      |||||:|||||:|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

318 DDP 320
51 TDP 53

RESULT 4
US-10-156-761-9468
Sequence 9468, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA

TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156, 761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 9468
LENGTH: 545
TYPE: PRT
ORGANISM: Streptomyces avermitilis
IS-10-156-761-9468

Query Match	4.4%;	Score 129;	DB 15;	Length 545;
Best Local Similarity	19.2%;	Pred. NO. 0.0023;		
Matches 100;	Conservative 57;	Mismatches 188;	Indels 176;	Gaps 19;
126	DKEWNELYSAEERLIGTSTK	-----EPDESIRH-----	TLVRLSLQDAYK	165
84	DKGNQOFFPEVNVYVCGNTKFGAALFRLRPDPDGE	LRHHDGSPAWPLSYEELEPYTT	142	
166	DRQRIF-----	RFLPLACHRLKNAPYVFW-HSAE-----	NLFPH-----	SIY 201
143	QAEHLVLVHGRHGCDTEGPTSAQYAYPVQGEPIQL	SHDLEKQGLHPHLPIGVNLT	202	
202	NDKQKLLFTLLTNHRCCTLAUTGGVEKKIGA-----	AEVRN	238	

203	QDDRGRATHA----	SACIECRDVRGFPCLVGA	AKSAQAQICVDP	PALEHANVEMLTHADVR	258
239	L-----	LATRNSSQLDSYINAK	VYVLASG	AGIGNPQILYNSFGSLQVTRND	286
259	LDTDATGRSVTSV	WATVGDGAATVEFSADI	VVWACGAVNSAVLLLSA	-----	307
287	SLIPNLGRYITEQ	PMAFQCIVLREQE	FVDSVRDD--	PYGLPMWKEAVAQHI	AKNPDTALPI 344

345 PRDPEPQVTTPTFTEHPWH-----TQHRDAFSY-GAVGPEVD 382

Sequence 15, Application US/09111969
Patent No. US20020137896A1
GENERAL INFORMATION:
APPLICANT: Kawamura, Yukio; Morita, Akihiro; Izumo, Koji.; Saka, Tomohide.
TITLE OF INVENTION: ANTITUMOR PROTEIN AND GENE ENCODING SAME
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/911,969
FILING DATE: 24-Jul-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997
ATTORNEY/AGENT INFORMATION:
NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 35
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-911-969-15
Query Match 3.7%; Score 108; DB 10; Length 35;
Best Local Similarity 66.7%; Pred. No. 0.0031;
Matches 22; Conservative 2; Mismatches 9; Indels 0; Gaps 0;
Qy 453 SPPOFMDPLGLHLHLAGTTRIGFDKATTVADNNS 485
Db 2 SYPOFMAGLVHLHTGTTRIGTDQTSVADPTS 34
RESULT 7
US-09-820-843A-58
Sequence 58, Application US/09820843A
Publication No. US20030039963A1
GENERAL INFORMATION:
APPLICANT: Council of Scientific and Industrial Research
TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEIN
TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
FILE REFERENCE: Q63915
CURRENT APPLICATION NUMBER: US/09/820,843A
CURRENT FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 118
SOFTWARE: PatentIn version 3.0
SEQ ID NO 58
LENGTH: 775
TYPE: PRT
ORGANISM: C. pneumoniae CWL029
FEATURE:
NAME/KEY: misc_feature

OTHER INFORMATION: hypothetical protein
NAME/KEY: misc_feature
OTHER INFORMATION: gi|4376756
US-09-820-843A-58
Query Match 3.7%; Score 106.5; DB 11; Length 775;
Best Local Similarity 20.9%; Pred. No. 0.62;
Matches 84; Conservative 59; Mismatches 158; Indels 101; Gaps 20;
Qy 116 PKLSNDPAEDDKENNELYSEAEERLIGTSTKEPDE-----SIRHTVLVRLQDA-----Y 164
Db 443 PRRNGSPREDSPLMNALVGVAAHK-HGAKTKESSESSTPEISISAPIVRGWSQSSVSFIV 501
Qy 165 KDRQRIREFPLPLACHRLKNAPEYVWHSANLPHSYNDKQKLLFLLTNHRCRLALT 224
Db 502 MEDDHIFVDVPRKGGIYDVPSSPRMSPALEEDVFGD----- 540
Qy 225 GGYEKKIGAAEVRNLLATRNPSQSDSYIMAKYVIVLASGAIGNPQILYNSGFGSLQVTPR 284
Db 541 --YEVPI TSAE-----PSKDKNIYMTPR---LATPAIYD--LPSRFGSGSSRSRPS 584
Qy 285 ND---SLIENLGRYTEQPMFCQIVLRQEFVDSVRDDPYGLPFWKEAQAHTAKNPTDA 341
Db 585 SDVRSSSPN-RRGVPLPPVP--SPAWSEE--GSIYEDMSGAGAGESDYEDMSRSPS-- 637
Qy 342 LPIPRDPEPOVTPPTTEHPWHTQIHRDAF-----SYGAVGPPEVDSRVIVDLRFMGATDP 397
Db 638 ---PRGDLDEPIYANTPDNPP-TQNRIDRLLOERSGGASASPVPE-IYDEIPIWHGRPP 692
Qy 398 EANNLLVFQNDVQDGYSMFOP-----TFVRVPSTASNVRRKMMADMCEVASNLGGYL 450
Db 693 A-----TLPRPENTLNVSLRVSPGFGPEVRAALLSESVAVMVEAESIV 737
Qy 451 PTSPP-----QFMDP--GLALHLAGTTRIGFDKATTVADNNS 485
Db 738 PPTPEPGDGESEYLEPLGG-----LVATTKILLQKQWPRGESNA 775
RESULT 8
US-10-253-971-3
Sequence 3, Application US/10253971
Publication No. US20030070192A1
GENERAL INFORMATION:
APPLICANT: GEORGES, FAWZY
APPLICANT: DONG, JIN-ZHUO
APPLICANT: KELLER, WILF
APPLICANT: HUSSAIN, ATTA A. K.
APPLICANT: SELVARAJ, GOPALAN
APPLICANT: DATLA, RAJU
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR MODULATING SECONDARY PLANT
TITLE OF INVENTION: METABOLITES
FILE REFERENCE: 73776-159D
CURRENT APPLICATION NUMBER: US/10/253,971
CURRENT FILING DATE: 2002-09-25
PRIOR APPLICATION NUMBER: US 60/072156
PRIOR FILING DATE: 1998-01-22
PRIOR APPLICATION NUMBER: US 09/012453
PRIOR FILING DATE: 1998-01-23
NUMBER OF SEQ ID NOS: 7
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 546
TYPE: PRT
ORGANISM: Arthrobacter pascens
US-10-253-971-3
Query Match 3.7%; Score 106; DB 15; Length 546;
Best Local Similarity 20.6%; Pred. No. 0.39;
Matches 72; Conservative 45; Mismatches 131; Indels 102; Gaps 16;
Qy 210 FTLLTNHRCRLALTGGYEKKIGAAEVRNLL--ATRNPSQSDSYIMAKYVIVLASGAIGN 267
Db 224 FTLLTGLRARQLVFDA--DKRCTGVDVDSAFGRTHLSARCE-----VILSTGAIDS 274

QY 268 POILYNSGFS-----GLQVTPRNDLSLIPNLGRYTEQPMARCOIVLRQEFVDSVRDD 319
DB 275 PKLLMISGIGPAHLAEHGVLEVIV--DS--PGVGEHLQDHPGVVQPEAKQOMVOT----- 326
QY 320 PYGLPMWKEAV-----AQHAKNPTDALPIPRDP-----EPOVTPPT 358
DB 327 --STQWWEIGIFTPTEGLDRPDLMMHGVSPFDMTLRGVYPTENGFSLTPNVT----- 380
QY 359 BEHPWHTQIHRDAFSYQAV---GPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQDGYSM 415
DB 381 -----HARSRTVLRSRDRFDRKPAVDPRYE--TDPEGHDMRWVMVAGIR----- 422
QY 416 POPTFRYRPSTASNRARKMADN-----CEVASNLGGVLPSPQFMDPGLALH 465
DB 423 -----KAREIAQPAWBTGRELSPGTEAQTDEELQDIYRKT-----HNTVYH 466
QY 466 LAGTTRIG-PDKATTVADNNSLVDFANLYVAGNGTIRTGFGENPTLTSM 514
DB 467 PVGTVMGPADDDMSPLDPELRVKGVTGLRVADASVMPEHVTVPNPTVM 516

RESULT 9
US-10-156-761-9160
; Sequence 9160, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 9160
; LENGTH: 514
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-9160

Query Match 3.6%; Score 104; DB 15; Length 514;
Best Local Similarity 20.3%; Pred. No. 0.56;
Matches 92; Conservative 52; Mismatches 183; Indels 126; Gaps 20;

QY 152 RHTLVLSQDAYKDRQRIFRPLPLACHRLKNAPYVEWHIS-----AENLFHSIYNDKQ 206
DB 125 RHRLQNRIVPVADKDRNFIARDFVTAASALGVPVIEDFNARPPAEGAGFFSLAYEPESN 184
QY 207 KK-----LFTLLNHRCTRRLALTGYY-----EKKIGAAEVRNLLATRNPSQLDSVI 253
DB 185 RRSASVAYLHPVLDRLPNLTLRETWAYGLLPDGGRLTRVQVROSDGT-----TATVR 238
QY 254 MAKVYVLASGAIGNPQILYNSGFSGLQ-----VTFRNDSLIPNLGRYTEQPMARCOIV 307
DB 239 AARDMLLCAGAIIDTFRLLLSGVGPAQLRDLGIEVRAD--VPGVGENLLDHPES---VI 293
QY 308 LRQEF-----VDSVRDDPYGLRWKKEAVQAQHNKPTDALPIPRDPEQV-----TTPT 358
DB 294 VMEHTGFLPPNSAMDSADGL-----FLRRAP-----RGRPDLMFHFYQVPPFT 336
QY 359 EE-----HPWHTQIHRDAFSYG-AVGPEVDSRVIVDLRWFGATDPEANNLLVFQNDVQD 412
DB 337 VNTERLGHF-----AIHGVCMCTPNVPARSVGRMWLADADPTR----- 375

QY 413 YSMQPTFRYRPSTASNRARKMADNCEVASNLGGYLP-----TSPQFMDPGL 462
DB 376 -----QPCLOFRYFTDEGYDEKTIIVDGLGIAREVAAAAPLURDLKREVAPEGPDRTDAEL 431
QY 463 A-----LHACTTRIGFPDKATTVADNNSLVDFANLYVAGNGTIR-----TGGEN 508
DB 432 SAVGRKAANTVYHPAGTCRMG-----ADDDPLAVLDPQLRVGVSGVRVWDSLFPTM 484
QY 509 PTLTSMCHAIKARSIIINTLKGGTGDKNTGEHR 541
DB 485 PTINPMLTVLLAAERAVDLIKG---TRQPGEAR 514

RESULT 10
US-10-156-761-14017
; Sequence 14017, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 14017
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-14017

Query Match 3.6%; Score 103.5; DB 15; Length 342;
Best Local Similarity 23.7%; Pred. No. 0.33;
Matches 89; Conservative 41; Mismatches 109; Indels 137; Gaps 23;

QY 191 HSAENL-----FHSIYNDKQKXLFLLTNHRCTRRLALTGCGYKKGKGAAB-VRNLLAT 242
DB 40 HAVESLGTARAGGMHLAYVTNNALRTPDAVADH-LTELGIPTGPSDVITSQAQAVARLISE 98
QY 243 RNPSSQLDSYIMAKVYVLASGAIGNPQILYNSGFSGLQVTPRNDLSLIPNLGRYTEQPM 302
DB 99 QVPAG-----ARVLVI-----GGEGLRVLRERGLEP----- 125
QY 303 FCQIVLRQEFVDSVRDDP-----YG---LWKEAVAQH-IAK-----NPTDALPIP- 345
DB 126 -----VDSAEDDPAAVVQGVGGPELWGRFAEASVIAARGVPWFASNTD-LTIPS 174
QY 346 -----FRDEPQVT-TPFTEHPWHTQIHRDA-FSYGAVGP-EVDSRV 385
DB 175 ARGAPNGAAVEWRIATGAEPQVAGKPL-----PPMRETILRTGAERPLVVGDR 227
QY 386 IVDLRWFGATDPEANNLLVFQNDVQDGYSMQPTFRYRPSTASNRARKMADNCEVASN 445
DB 228 DTDLE--GAFNGEVDLSLLV--TGVTDGAQLLAAPRRHPT-----YVDAD 269
QY 446 LGGYLPSPQFMDPGLALHLAGTTRIGFDRKATTVAONNSLVDFANLYVAGNGTIRTF 505
DB 270 LRGLL-TQQPEWEGDGFRCGGWT-----ATAGSE-----RLELAGEAMDG- 312
QY 506 GENTLTSMCHAIKSA 521
DB 313 -----LRALCAAAWTA 323

RESULT 11

US-09-911-969-6
; Sequence 6, Application US/09911969
; Patent No. US20020137896A1
; GENERAL INFORMATION:
; APPLICANT: Kawamura, Yukio; Morita,
; Akihiko; Izumo, Koji.; Saka, Tomohide.
; TITLE OF INVENTION: ANTITUMOR PROTEIN AND
; GENE ENCODING SAME
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 PARK AVENUE
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154

COMPUTER READABLE FORM:

MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: MICROSOFT WORD 97

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/911.969
FILING DATE: 24-Jul-2001

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/023,731
FILING DATE: 13-FEB-1998
APPLICATION NUMBER: JP 29275/1997
FILING DATE: 13-FEB-1997

ATTORNEY/AGENT INFORMATION:

NAME: KENNETH H. SONNENFELD
REGISTRATION NUMBER: 33,285
REFERENCE/DOCKET NUMBER: 3479-4000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 758-4800
TELEFAX: (212) 751-6849
TELEX: 421792

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 61
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:

US-09-911-969-6

Query Match 3.5%; Score 102.5; DB 10; Length 61;
Best Local Similarity 43.5%; Pred. No. 0.026;
Matches 27; Conservative 6; Mismatches 24; Indels 5; Gaps 2;
QY 29 PNVVIGALQVSVPRNQNVPTLDPGAWSPGSSA-----ISNGKNPHOREPENSAAE 84
Db 1 PNVVINGALQVSVPRNQNVPTLDPGAWSPGSSA-----ISNGKNPHOREPENSAAE 84
QY 85 VT 86
Db 60 VT 61

RESULT 12

US-10-092-219-2
; Sequence 2, Application US/10092219
; Publication No. US20020115114A1
; GENERAL INFORMATION:
; APPLICANT: Domin, Jan
; TITLE OF INVENTION: No. US20020115114A1el Lipid Kinase
; FILE REFERENCE: 1064HG/50947
; CURRENT APPLICATION NUMBER: US/10/092,219
; CURRENT FILING DATE: 2002-03-07
; PRIOR APPLICATION NUMBER: PCT/GB98/00244
; PRIOR FILING DATE: 1998-01-27
; PRIOR APPLICATION NUMBER: 09/355,160

; PRIOR FILING DATE: 1999-10-01
; PRIOR APPLICATION NUMBER: 9701652.1
; PRIOR FILING DATE: 1997-01-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 1686
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-092-219-2

Query Match 3.5%; Score 102.5; DB 14; Length 1686;
Best Local Similarity 24.3%; Pred. No. 5.2; Indels 73; Gaps 19;
Matches 74; Conservative 41; Mismatches 116;
QY 50 PTLDPGAWSPGSSA-----SSAISNGKNPHOREPENSAAEAVTRGVG 90
Db 135 PTLDPGAWSPGSSA-----SSAISNGKNPHOREPENSAAEAVTRGVG 190
QY 91 GMSHTW-----TCSTPRIHPMESLPGIRPKLSNDPAEDDKENNELYSEARL-----IGT 142
Db 191 GQSPYFSLPTATP-PH-PQGSLLP-IYRPVSTDMA---KLFDKIATSEFLKNGKART 244
QY 143 STKEFDESIRHLVLRSLQDAYKDRORIFRPL---PLACHRLKNAEYVWHSAENLFHS 199
Db 245 DLEITDSKSVNLQVSPKSEDISK-----FDWLDLDPKPKVDNV-EVLDPHEEKNVSSL 298
QY 200 IYNDKQKFLTLTNHRCRLALTGGEYKKGICAAEVRNLLATRNPSQLDSYIMAKVYV 259
Db 299 LAKDPWDVLLBERSTANC-----HLERKVNKSLSVATVTRSSQSLNRTTQLAK--- 348
QY 260 LASGAIGNPQILYNSGFGSLQVTPRNDLSLPNLGRVITEQPMACQIV--LRQEFVDSVR 317
Db 349 -ACGHISQKD---PNCSTSL---PTGSSLLQEV-EVQNEEMAAFCRSITKLKTKF----- 395
QY 318 DDY 321
Db 396 --PY 397

RESULT 13

US-09-815-915-8
; Sequence 8, Application US/09815915
; Patent No. US20020025931A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Meyers, Rachel
; TITLE OF INVENTION: 3714, 16742, 23546, AND 13887 NOVEL
; TITLE OF INVENTION: PROTEIN KINASE MOLECULES AND USES THEREFOR
; FILE REFERENCE: 38155-20006.00
; CURRENT APPLICATION NUMBER: US/09/815,915
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 60/191,846
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 1244
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-815-915-8

Query Match 3.5%; Score 102; DB 9; Length 1244;
Best Local Similarity 17.4%; Pred. No. 3.6; Indels 190; Gaps 23;
Matches 95; Conservative 65; Mismatches 195;
QY 34 KGALQVSVPRNQNVPTLDPGAWSPGSSAISNGKNPHOREP-----ENL 80
Db 661 EGFLTAITIP-----RPSVASTQSTSGSPHCQOPEKDLQPMPTVELYSRENPF 711
QY 81 SAEAVTRG---VCGMSTHWTCTSTPRIHPNMESLPGIRPKLSNDPAEDD-----KE 128
Db 712 SGLVVTGEPPSGSRTDLGLQIDHIGHDN--LPNRESNKSQDLGPKELPDHNRVRE 769

QY 129 WNELYSEA-ERLIGTSTKEFDESI---RHTLVLSLOD---AVKDRQIRFRPLPLACHR 180
Db 770 FENLPGETEEKSILLESNDEKLSRGQHCIEISLPGDLVIVEKDHSAATEPLDV---825
QY 181 LKNAPEVEVHSAENLFHSIYNDKQKLLFTLLNHRCTRLALTGGYEKIGAAEVRNLL 240
Db 826 ---TKTQTFSVVFNQDKNEIMKLLT-----VGTSEI---854
QY 241 ATRNPSQLDSYI---MAKVYVLASGAIGNPQILYNSGFGSLQ-----VT 282
Db 855 ---SRDIDPHVEGQIGQVAEMQKNKISKDDIMSEDLPGHOGDLSTFLHQBKREKIT 910
QY 283 PRNDSL-----IPNLGRYITE-----298
Db 911 PRNGELFHCVSENEHGAPTRKDMVRSSFVTRHSRIPVLAQEIDSTLESSPVSACEKLLQ 970
QY 299 ---QPMACQIVLROEFVDSVRDDPYGLPMWKEAVAQHIKNTDALTPIPRDRPQVPT 354
Db 971 KKAYQPDVLKLVKQKFSFLGD-----LSSASDKLEEKLATVPAPF---CEEVL 1020
QY 355 TPPT-----BEHPWHTQIHRDAPSYGAVGVEVDSRVIVDLRWFGATDPEANLLVFQNDVQ 410
Db 1021 TFSRLTVDSH---LSSAEDSFLSPIISQSRKSKIPRPVSWVNTDQVNSTSQF-----1073
QY 411 DGYSMPQPTFRYRPSTASNVRAR-----KQMDMCEVASNLGGYLPSPPO 456
Db 1074 ---FPRPPPG---KPTRPGVEARLRRYKVLGSSNSDSDFSLAQILON-GSQKPRSTTQ 1127
QY 457 FMDPG 461
Db 1128 CKSPG 1132

RESULT 14
US-10-393-316-8
Sequence 8, Application US/10393316
Publication No. US20030175786A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 3714, 16742, 23546, AND 13887 NOVEL
FILE REFERENCE: MP100-086P1RCNIM
CURRENT APPLICATION NUMBER: US/10/393,316
CURRENT FILING DATE: 2003-03-20
PRIOR APPLICATION NUMBER: US 09/815,915
PRIOR FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: US 60/191,846
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 19
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 1244
TYPE: PRT
ORGANISM: Homo sapiens
US-10-393-316-8
Query Match 3.5%; Score 102; DB 12; Length 1244;
Best Local Similarity 17.4%; Pred. No. 3 6;
Matches 95; Conservative 65; Mismatches 195; Indels 190; Gaps 23;
Y 34 KGAQQSVVVRNQNVPLDQWAPPGSSAISKPHQREF-----ENL 80
b 661 EGPLTAITIP-----RPSVASTQSTSGFHCQKQPEKDLQPMETVELYSPRENF 711
Y 81 SAEAVTRG---VGGNSTHTWCTSPRIHPMESLPGIGRPKLSNDPAEDD-----KE 128
b 712 SGLVTEGPEPPSGSRTDLQIDHIGDM---LPNIRESNKSQDLGPKELPDHNLVVRE 769
Y 129 WNELYSEA-ERLIGTSTKEFDESI---RHTLVLSLOD---AVKDRQIRFRPLPLACHR 180
b 770 FENLPGETEEKSILLESNDEKLSRGQHCIEISLPGDLVIVEKDHSAATEPLDV---825

QY 181 LKNAPEVEVHSAENLFHSIYNDKQKLLFTLLNHRCTRLALTGGYEKIGAAEVRNLL 240
Db 826 ---TKTQTFSVVFNQDKNEIMKLLT-----VGTSEI---854
QY 241 ATRNPSQLDSYI---MAKVYVLASGAIGNPQILYNSGFGSLQ-----VT 282
Db 855 ---SRDIDPHVEGQIGQVAEMQKNKISKDDIMSEDLPGHOGDLSTFLHQBKREKIT 910
QY 283 PRNDSL-----IPNLGRYITE-----298
Db 911 PRNGELFHCVSENEHGAPTRKDMVRSSFVTRHSRIPVLAQEIDSTLESSPVSACEKLLQ 970
QY 299 ---QPMACQIVLROEFVDSVRDDPYGLPMWKEAVAQHIKNTDALTPIPRDRPQVPT 354
Db 971 KKAYQPDVLKLVKQKFSFLGD-----LSSASDKLEEKLATVPAPF---CEEVL 1020
QY 355 TPPT-----BEHPWHTQIHRDAPSYGAVGVEVDSRVIVDLRWFGATDPEANLLVFQNDVQ 410
Db 1021 TFSRLTVDSH---LSSAEDSFLSPIISQSRKSKIPRPVSWVNTDQVNSTSQF-----1073
QY 411 DGYSMPQPTFRYRPSTASNVRAR-----KQMDMCEVASNLGGYLPSPPO 456
Db 1074 ---FPRPPPG---KPTRPGVEARLRRYKVLGSSNSDSDFSLAQILON-GSQKPRSTTQ 1127
QY 457 FMDPG 461
Db 1128 CKSPG 1132

RESULT 15
US-09-801-368-322
Sequence 322, Application US/09801368
Patent No. US20020128250A1
GENERAL INFORMATION:
APPLICANT: Busby, Robert
APPLICANT: Call, Brian
APPLICANT: Hecht, Peter
APPLICANT: Holtzman, Doug
APPLICANT: Madden, Kevin
APPLICANT: Maxon, Mary
APPLICANT: Milne, Todd
APPLICANT: No. US20020128250A1man, Thea
APPLICANT: Royer, John
APPLICANT: Salama, Sofie
APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801.368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1999-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 322
LENGTH: 1536
TYPE: PRT
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-322

Query Match 3.4%; Score 97.5; DB 10; Length 1536;
Best Local Similarity 19.4%; Pred. No. 14;
Matches 103; Conservative 71; Mismatches 163; Indels 193; Gaps 28;
QY 58 SAPGSSAISKPHQREFENLSA-----EAVTRGVGGNSTHTWCTSPRIHPMESLP 111
Db 1044 GSGDDGSSIASRKPKPYQEMSLLDILHRSYQKLR-----SNDEKVPOLSEPPSEEP 1098
QY 112 G-IGRPKLSNDPAEDD-----KEMNELYSEA 136

```
Db      1099 NTIEEBELIDBEAKPWLTGNLVEANSQGIIONRSIFNLPANTNVIYFPRHWTIY--- 1155
Qy      137 ERLIGTSTKEPDESIRHTLVRLS-----QDAYKDRQIRFRP 173
Db      1156 ERL--EIKQNNERTKINTRESTVTFPAKDLDDLSSQSEMGDFVGEDAYKQVRLSR- 1212
Qy      174 LPLACHRLKNAPEYVEVHSAENLFHSIYNDKQKLF-----LTNHRCTRLALTG 227
Db      1213 -----RLINGDLEHWF--EESLRQAYN-NKAFKLYTIDKVTQSLVKHAHT--LMTDAK 1261
Qy      228 EKKIGAAEVRNLLATRNPSSQLDSYIMAKVYVVLASGAIGNPQIILYNSGFSGLVTPRND 287
Db      1262 TAEIMALFVKD-----RN-----ASTTSAKDQIIYR-----LQVR-SHMS 1295
Qy      288 LIPNLGR-----YITEQPMACQIVLROEFVDSVRDDPYGLPWNKEAVAQHIKNPT 339
Db      1296 NTENMFRIEFKRTLHVSIQYIALDDLTLKEPKADEK-----WKYVYTSYALPHPT 1347
Qy      340 DALPIPPRDPFPQVTPTEHPHWTQIHRDAFSYGAVGPEVDSRVIVDLRMFCATDPEA 399
Db      1348 EGI-----PHEKLKIPLE-----RLIEFGQDIDGTEVDEEF-----SPEG 1383
Qy      400>NNL-----LVFONDVQDQ-YSM--PQTPRYRPTASNVPRKXMMADMCEVASNLG 447
Db      1384 ISVSTLKIKIQPITYQLHIENGSDYVTRKATNKY-PTIANDNTQKGWVQKKELIS--- 1439
Qy      448 GYLTPSPQFMD--PGLALHLAGTTRIGFDK-----ATTVADNNSL 486
Db      1440 -----KFLDCAVGLNLDLDEAKLSMQKKNLXDSIAKTSAGNOGI 1481
```

Search completed: October 28, 2003, 16:09:44

Job time : 74 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model
Run on: October 28, 2003, 15:54:42 ; Search time 42 Seconds
(without alignments)
1243.324 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618
Perfect score: 2901
Sequence: 1 NABEGTAVPVPGVHKQNEI.....IINTLKGTGDKNTGERNL 543

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues
Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:
1: PIR1:
2: PIR2:
3: PIR3:
4: PIR4:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match Length	ID	Description
1	170.5	5.9	573	2 E81385
2	166.5	5.7	579	2 H87451
3	159	5.5	615	2 B38575
4	158.5	5.5	529	2 A84260
5	158.5	5.5	619	2 AB2003
6	144	5.0	722	2 B75608
7	141	4.9	591	2 H83362
8	140.5	4.8	562	2 JC7628
9	135	4.7	561	2 AE3094
10	135	4.7	561	2 E98192
11	128.5	4.4	748	2 T10651
12	122	4.2	527	2 AD3082
13	122	4.2	527	2 E98204
14	116	4.0	502	2 T18562
15	116	4.0	514	2 B70917
16	116	4.0	599	2 T19711
17	114	3.9	1012	2 T00958
18	113.5	3.9	578	2 F70736
19	111.5	3.8	494	2 AD3255
20	110.5	3.8	589	2 T50698
21	110	3.8	1498	2 AF108
22	108	3.7	1125	2 T12193
23	106.5	3.7	775	2 B72074
24	106.5	3.7	775	2 C81594
25	106.5	3.7	775	2 D86549
26	105.5	3.6	1611	2 T38236
27	105	3.6	499	2 T45749
28	104	3.6	1917	2 C88728
29	102.5	3.5	768	2 JC6564

30	102	3.5	1018	2 T40253
31	102	3.5	1042	2 T16169
32	102	3.5	2361	2 T25752
33	101.5	3.5	489	2 C82241
34	101.5	3.5	539	2 C81805
35	101.5	3.5	747	2 F87283
36	100.5	3.5	517	2 G81060
37	100	3.4	1142	2 T37455
38	99.5	3.4	501	2 B81048
39	98.5	3.4	770	2 S60676
40	98	3.4	625	2 T20634
41	98	3.4	639	2 T43190
42	98	3.4	7576	2 T17428
43	97.5	3.4	328	2 AF0170
44	97.5	3.4	569	2 E86957
45	97.5	3.4	578	2 D95400

ALIGNMENTS

RESULT 1

E81385
Probable oxidoreductase chain Cj0415 [imported] - Campylobacter jejuni (strain NCTC 1116)
C/Species: Campylobacter jejuni
C/Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 03-Jun-2002
C/Accession: E81385
R/Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillin:
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; Barrell:
Nature 403, 685-688, 2000
A/Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hyp:
A/Reference number: A81250; MUID:20150912; PMID:10688204
A/Accession: E81385
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-573 <PAR>
A/Cross-references: GB:AL139075; GB:AL111168; NID:G6967817; PIDN:CAB74251.1; PID:G696788:
A/Experimental source: serotype O2, strain NCTC 11168
C/Genetics:
A/Genes: Cj0415

Query Match	5.9%	Score 170.5;	DB 2;	Length 573;
Best Local Similarity	20.9%	Pred. No. 4.9e-05;		
Matches 111;	Conservative 61;	Mismatches 237;	Indels 123;	Gaps 21;
QY	77	FENLSAEAVTRGVGGHSTHTCTSTPRIHPMSLPGIRP-----KLSNDPAEDD-----K	127	
Db	84	YRMGSFLIGNNVGGAGVHNGWTFPMFYDFEIQTLKQRYGNKLGNDYTLQDWGVYK	143	
QY	128	EWNELYSEABRLIGTSTKEF-----DESIRHTLVLSLQDAYKDRQIFR	172	
Db	144	DMEPYDREKTCGVSGEENPLAEKMGAFRSPYPOEPLNTMLKRFESAASSNLHTY	203	
QY	173	PLPLA-----CHRLKNAPEYVHSAENLFHSIYNDKQKCLF	210	
Db	204	RLPANSKGGYTPDQDLAPQCYCAYCERFG-----CEYGAKASPLNTVIPKANSTGKY	258	
QY	211	TLLTWHRCTRLALTGGYKKGICAAEVNLLATRNSSQLDSYIM-AKVTVLASGAGNQPQ	269	
Db	259	TIRTSYNTVQIL-----KDGKVTGVKPFVDTRT-----MKEYIQPADIVLVTSYMFNNAK	308	
QY	270	ILYNSGFSGLQVTP-----RNDSLIPNLG--RYITEQPNAFQIVLRQBFVDSVRDD	319	
Db	309	LMVSNV-GEQVDPKTGKTGLGRNYCYQNNMGTAFDEQFNTP-----MGSGALGTTSD	363	
QY	320	PYG--LPWKE-----AVAQHAKNPTDALPIPRPEPQVTPFTTEHHPWHTQIHRD	370	
Db	364	FNGDNFDSKKEFLGAMTYSVQLGTRPIQSPAPAGAP-----TWGAEF-KK	410	
QY	371	APSYG-----AVGPEVDSRVIVDLRWFGATPEANLLVFQNDVQDGYSPMPQTFRYRPS	425	
Db	411	ALNVNFTRAITVGGQ-----GASLPHKNYLSLDPTYKDAFGMPLRLITYN-F	457	

R:Kaneke, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anai
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AB2003
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-619 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA077942.1; PID:gl7135396; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1576

Query Match 5.5%; Score 158.5; DB 2; Length 619;
Best Local Similarity 19.7%; Pred. No. 0.00046;
Matches 125; Conservative 72; Mismatches 217; Indels 221; Gaps 27;

QY 22 FQKIDRFVNVKALQOVSVVRNQVVTLPDGAWSAPPSSA-----INSGN 71
Db 60 FQSYVDTFYKAVD-----KNPNSP-----YPANSNVQSPDYNDYFIEQGPM 101
QY 72 PHOREPENLSAAVTRGVGSMTHWTCSTRIHP---PMESLPGIGRPKLSNDPAEDDKE 128
Db 102 P-----LAGSYTRVLGGTTMHWEAKTPRLPEDFKLSSTYVQGL-----DWPIDYHD 148
QY 129 WNELXSEARLIGTSTK-----EPDESIRHTLVLSLQDAYKDRQRI----- 170
Db 149 LEFYRKAHEHGVCGDVDEQALGLEFPQD---YFPMKLPSPYLDQKIEKVGNTVE 206
QY 171 -----FRLPLACHRLKNAPEYVHSAENLF-----HSIYNDDK----- 205
Db 207 LYGKTHLTSFSTFPOARNGVPN-PKY-----DQNLFPDGVTSVHPVOYGERCOGNACV 261
QY 206 -----QKGLFTLLTHRCTRLALTGCGYKKGIGAAEVNRNLLATRNPSQLD 250
Db 262 PICPVOAKYDARLTLSKAFETGKVLVQAVKVEYDQTRITAIHYKHKKPNSSEY 321
QY 251 SYIMAK---VYVLASGAIGNPOILVNSGFGSGLVTPRNDLSIPLNGLRYITEQPMACQIVL 308
Db 322 TTGIAGKTLFVLATNAVENARLLGSDL-----PNTSLI---GTYL----- 360
QY 309 ROEFVDSVRDDPYGLPWKEAVAQHIKKNPTDLPPIPRDPE-----PQVTPPF----- 357
Db 361 -----MDHEFTLAW-----ALM-----PEVTGTMRGPLVTSGIGTFR 392
QY 358 -----TESHPWHTQIHRDAFSYCAVGPEVDSRVVD-----LR 390
Db 393 KDFRKKQSAFAVDIHDGNGWATGSPKSEVEDAVDNKNYQGLRQLTISRQLLLA 452
QY 391 WFGATDPEANLLVFQNDVQDQYSGMPQPTFRYRPSSTASNV---RARKMADMCEVA---S 444
Db 453 FMCCELLPEYGNRVITDPRHKKLG-----NVRPVINENLDPYSRRTLAYTKVSRVPE 506
QY 445 NLGG-----YLTPSPQPMQDGLAL-----HLACTTRIGFDKATTVADNNSLVWDFANL 493
Db 507 RLGAEDYTHYDQDPAYEPEGEYVYKGNHFSGTHIMGTPLNSVVDYLSRWDHKNL 566
QY 494 YVAGNGTIRTGFGNPTLTSMCHAKSARSINTL 528
Db 567 FLVGAGSMPTIGSSNTTLTIALSPTAEHMLQEL 601

RESULT 6
B75608
GMC oxidoreductase - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: B75608
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; I
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999

564 SWDVNVFVFGASAPFQGLGYNPT 587
Db 564 SWDVNVFVFGASAPFQGLGYNPT 587

RESULT 4
AB4260
hypothetical protein Vng1035c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: AB4260
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ehardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: AB4160; MUID:20504483; PMID:11016950
A:Accession: AB4260
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <STO>
A:Cross-references: GB:AE004437; NID:gl0580587; PIDN:AAG19445.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG1035C

Query Match 5.5%; Score 158.5; DB 2; Length 529;
Best Local Similarity 22.3%; Pred. No. 0.00036;
Matches 117; Conservative 44; Mismatches 194; Indels 169; Gaps 22;

QY 84 AVTRGVGGMSTHTWCTSTPRIHP---MESLPGIGRPKLSNDPAEDDKENNELYSEARLI 140
Db 88 ARVKGVGSTLHWQNMWMLHEQDFLASATGVA-----DWPFDVTLKPYAAAESAL 142
QY 141 -----GSTKEFDESIRHTLVLSLQDAYKDRQRIFRPLPLACHRLKNAPEYVHSAENLF 197
Db 143 GYSGASDNPFAFPPEQHPQAPPPSYSD---SLFAD---ACESLGIAT----- 185
QY 198 HSIYNDDKQKLLFTLLTHRCTRLALTG-----GYE-----KKIG 232
Db 186 HSPVN-----ARLSAGRETRACVGTGCPVCPGSAKYDATVHVDRATDAGARVID 237
QY 233 AAEVRNL-----LATRNPSSQLDSYIMAKVYVLASGAIGNPOIL-----Y 272
Db 238 EAPVQLEHDAAGDRVTGAVATPDGTHRQS---ATEFVLAAGGLETPELLLSDSRY 294
QY 273 NSGF---SGLQVTPRNDLSIPLNGLRYITEQPMACQIVLRQBFVDSVRDDPYGLPWKEA 329
Db 295 PDGLANSSGL-----VGRY-----FMD----- 311
QY 330 VAQHIKAKNPTDALPIPRDPEPQVTPPTTEHPWHTQIHRDAFSYCAV-----GPEV 381
Db 312 ---HLFAGAGGTLDEFTQRNHYGFNT---TESHQY---DRPDGSRGAIKLEFLNAGPSP 363
QY 382 DSRVVDLRWFCA-----TQPEANLLVFQNDVQDQYSGMPQPTFR 421
Db 364 REMALSGDDWDGMDCDRIDASGTHIAUGLVQEQPRPENRVLHPERTDVHGNVPDQV 423
QY 422 YRPSTASNVRAKMMADMCEVASNLGGYLP-TSPQFMDPGLALHAGTTRIGFDKATTV 480
Db 424 WLSLAYERRTIERANIEQREILLTALGADIEMTVGPB---DTGPAFHHMGTTRNGTDPAESV 481
QY 481 ADNNSLVWDFANLYVAGNGTIRTGFGNPTLTSMCHAKSARS 524
Db 482 VDPRLRTHLSNLSVASSSVFPTAGAMNPTLTIALAKAADI 525

RESULT 5
B2003
hydrogenase chain [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp. PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Dec-2002
C:Accession: AB2003

Best Local Similarity 23.7%, Pred. No. 0.0097;
Matches 45; Conservative 29; Mismatches 77; Indels 39; Gaps 5;
QY 372 PSY-GAVGPEVDSRVIVDLR-----W-----FGATDEANNLLVFQNDVQ 410
Db 379 FGVOGAASREGWDREIAELNIGADLKQALTOPGGWTIGTGMFLPDHNRISLDHSVR 438
QY 411 DGYSMPQPTFRYRSTASNRVRKRWADMCE-----VASNLGGVLPSPQFWDPG 461
Db 439 DKWGLPVLSDIDVELKQNERDVRDMVDVADLLEAAGVKNVKGVDGYA-----PG 489
QY 462 LAHLAAGTTRIGFDKATTVADNNSLVWDFANLYVAGNGTTRTGFGENPTLTSCHAIKSA 521
Db 490 MGIHENGTMGRDPTKTSVLNSHQVWDAPNVFTDGCACHTSSSCVNPSTLYMALTARAV 549
QY 522 RSIINTLKGK 531
Db 550 DYAVEELKRG 559
RESULT 9
AE3094
Oxidoreductase Atu4377 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 18-Nov-2002
C:Accession: AE3094
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.;
Karp, P.; Romero, P.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClellan,
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; MUID:21608550; PMID:11743193
A:Accession: AE3094
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE008699; PIN:AA45171.1; PID:g17742847; GSPDB:GN00187
A:Experimental source: strain C58 (Dupont)
A:Genetics:
A:Gene: Atu4377
A:Map position: linear chromosome
Query Match 4.7%; Score 135; DB 2; Length 561;
Best Local Similarity 24.8%; Pred. No. 0.026;
Matches 39; Conservative 25; Mismatches 59; Indels 34; Gaps 3;
Y 392 FGATDEANNLLVFQNDVQGYSMPOPTFRYRSTASNRVRKRWADMCE----- 441
b 420 FGEMLPYHDNRVKLDHDKKDWGL-----PVLSSNVEMKQNELDMEDMVDVAVEMF 471
Y 442 -----VASNLGGVLPSPQFMDPGLALHLAGTTRIGFDKATTVADNNSLVWDFANLY 494
b 472 EAVGIKNVKSRTYA-----PGMGIHENGTMGRDPTKTSVLNGNQNVDAPNVF 522
Y 495 VAGNTGTRTGFGENPTLTSCHAIKSAISINTLKGK 531
b 523 VTDGACMTSASCNVPSLTLYMALTARAAEFASVSRKKG 559
RESULT 10
E98192
Probable oxidoreductase chain C0415 [imported] - Agrobacterium tumefaciens (strain C58,
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 18-Nov-2002
C:Accession: E98192
A:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Woliam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: E98192

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-561 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK89063.1; PID:g15158859; GSPDB:GN00170
C:Genetics:
A:Gene: AGR L 980
A:Map position: linear chromosome
Query Match 4.7%; Score 135; DB 2; Length 561;
Best Local Similarity 24.8%; Pred. No. 0.026;
Matches 39; Conservative 25; Mismatches 59; Indels 34; Gaps 3;
QY 392 FGATDEANNLLVFQNDVQGYSMPOPTFRYRSTASNRVRKRWADMCE----- 441
Db 420 FGEMLPYHDNRVKLDHDKKDWGL-----PVLSSNVEMKQNELDMEDMVDVAVEMF 471
QY 442 -----VASNLGGVLPSPQFMDPGLALHLAGTTRIGFDKATTVADNNSLVWDFANLY 494
Db 472 EAVGIKNVKSRTYA-----PGMGIHENGTMGRDPTKTSVLNGNQNVDAPNVF 522
QY 495 VAGNTGTRTGFGENPTLTSCHAIKSAISINTLKGK 531
Db 523 VTDGACMTSASCNVPSLTLYMALTARAAEFASVSRKKG 559
RESULT 11
T10651
Hypothetical protein TSP17.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 15-Sep-2000
C:Accession: T10651
R:Bevan, M.; Lennard, N.; Quail, M.; Harris, B.; Rajandream, M.A.; Barrell, B.G.; Bancroft,
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z16533
A:Accession: T10651
A:Molecule type: DNA
A:Residues: 1-748 <BEV>
A:Cross-references: EMBL:AL049917; GSPDB:GN00062; ATSP:T5F17.20
A:Experimental source: cultivar Columbia; BAC clone T5F17
C:Genetics:
A:Gene: ATSP-T5F17.20
A:Map position: 4
A:Introns: 70/3
C:Superfamily: Arabidopsis thaliana hypothetical protein T5K18.160
Query Match 4.4%; Score 128.5; DB 2; Length 748;
Best Local Similarity 19.7%; Pred. No. 0.13;
Matches 105; Conservative 77; Mismatches 187; Indels 165; Gaps 24;
QY 76 EPNLSAEAVTRGVGCMSTHTCTSPRIHPPMESLPGIGRPKLSNDPAEDDKEMNELYSE 135
Db 301 KFMLLAGSAGV---GGTAVNWSAS---TTPDHLV-----QEW-----SE 334
QY 136 AERLIGTKSF---DE-SIRHTLVLSLDQAYKDRQRIPLPLACHRL-----K 182
Db 335 GSKIFFGSEQYQSAWDEVTIRIGVTCVKHGFQV---QVLAKGCEGLQGVESVPR 389
QY 183 NAPE-----YVEMHSAENLFHSIYNDKQKLFLLTNHRCRLAL-----TGGEYKK 230
Db 390 NSPDHYCGLCGYCRAGAKNGTDQTLWLDVAVENGAVILTGKAERFVLVDNVTSSNERK 449
QY 231 -----IGAAEVRNLIAITRNPSQLDSYIMAKYVLASGAIGNPQILYNSGFSGLQVTPRN 285
Db 450 KRCVGFVASSVGGKIGKK-----FIEARVTSSAGSLTTPMLSSGLKN----- 495
QY 286 DSLIPNLR-----YITEQMAFCQIVLRQBFVDSVR--DDPYGLPMMWEAVA 331
Db 496 ---PNIGRNKLHPVLMTCWYFFEKDSEFGKMGYEGGIITSVHMND-----TESGC 544
QY 332 QHIAKNPTDALPIPRDPEPOVTPPTTEHPWHQIHRDAFSYAGVGEVDSRVIVDLRW 391
Db 545 KAILENPLIG-PASYAGLSPPWVS-----GDLKERMIK----- 576

submitted to the EMBL Data Library, August 1997

A;Accession: Z18976

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-502 <WIL>

A;Cross-references: EMBL:Z98262; PIDN:CAB10932.1; GSPDB:GN00019; CESP:F15C11.2

A;Experimental source: clone VF15C11L

R;Wilkinson, J.

submitted to the EMBL Data Library, April 1996

A;Reference number: Z19352

A;Accession: Z20370

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: DNA

A;Residues: 1-502 <WIL>

A;Cross-references: EMBL:Z71260; PIDN:CAA95799.1; GSPDB:GN00019; CESP:F15C11.2

A;Experimental source: clone F15C11

C;Genetics:

A;Gene: CESP:F15C11.2

A;Map position: 1

A;Introns: 31/3; 67/1; 169/3; 212/3; 293/2; 294/1; 340/3; 368/3; 406/2; 428/2; 448/3; 48

Query Match

Best Local Similarity 4.0%; Score 116; DB 2; Length 502;

Matches 110; Conservative 65; Mismatches 201; Indels 126; Gaps 26;

40 VSVVRNQVPTLDGAWSPGSSAISNGKPHQREFENLSAEAVTRGVGGMSTHTWCS 99

74 VHLVIRNQARPTPAPAA-ATPTASSAPSS--NP-----103

100 TPRIHP-----PMESLPGLIGRP-KLSNDP-----ABDDKEMNELYSEAR--RLIGTSTKE 146

104 TPSSQPNTPNPFPAAGMGSGFADILNPDAMRSVNDNITQQLGNPEFMTIQSNPQ 163

147 F-----DESIRHTL-----VLSLQDAYKDR---QRIFRPLPLACHRLKNAPEYVEVHSA 193

164 FOALIERNPVEVGHILNDPNVMQTMEMIRPNMFOEMMRNHQAIRNLQIP-----GG 217

194 ENLFHSIYNDKQKLFLLTNHRCITRLALTG-----GYEKKIGAAEVRLNLLATR 243

218 EAALERLYN-DVQEPFLNSATN-----SLSNPFFASLGDQSSFRVDRAGCENNEALP 270

244 NPSSQLDSYINAKVYVVLASGAIIGNPOILYNSGFGSLQVTPRNDLSLIPNLGRYITQPMAF 303

271 NPWAS-----NANQATNQSNNRSADFNLLDSFGISSLEQNMSPSQASMF 319

304 COIVLRQEFVDSVRDDPYGLPWWKEAVAQHIARNP--TDAL--PIPFRRPEPQVTTPTTE 359

320 -----SPEVINSIR-----QNMSSNPLDLSIVGQIPSDARDNPQISEGIRR 360

360 EHPWHTQTHRDAPSYGAV-GPEVDSRVIVDLRWFGATDPEANNLL-VFQNDVQDGYSMQP 417

361 SFPQMLNMWSDPSVMEAMRNPRVSEAFRIQEGFSTLRREAPQLNLNFOAGAMGGGAFGS 420

418 PTERYRPSSTASNRKMKWADCEVASNLGGVLPSPQFMDP--GLALHLATGTRIGF- 474

421 DA-----NASSAGANSGLADLFN-SMNMGGRPSSTAAPVNPEQTYASQLEQLQSMGFS 475

475 DKATTVADNNSLVMDFANLYVA 496

476 DRARNA---ALTATFGDLNAA 494

RESULT 15

70917

robable zwf2 protein - Mycobacterium tuberculosis (strain H37RV)

Species: Mycobacterium tuberculosis

Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000

Accession: B70917

Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

Connor, R.; Davies, R.; Feltwell, T.; Dentles, S.; Hamlin, N.; Holroyd, S.

ajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

ature 393, 537-544, 1998

A;Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrall, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome :
A;Reference number: A70500; MUID:98295987; PMID:9634230

A;Accession: B70917

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-514 <COL>

A;Cross-references: GB:Z98444; GB:AL123456; NID:g3250713; PIDN:CAB09259.1; PID:g2131049

A;Experimental source: strain H37RV

C;Genetics:

A;Gene: zwf2

C;Superfamily: glucose-6-phosphate dehydrogenase

Query Match

Best Local Similarity 4.0%; Score 116; DB 2; Length 514;

Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;

29 FVNVIKALQQ-VSVVVRNQVPTLDGAWSPGSSAISNGKPHQREFENLSAEAVTR 87

77 FQOVYNAVOEHCRTPFROQWDRLAEG-FRVPGTFDDDDAFQAETLEKLDAAE---R 132

88 GVGGMSTHTWCTSTPRIHPPM-ESL--PGIGRPKLSNDPAEDDKEMNELYSEARLIGTST 144

133 GTGGNHAFYLAIPPKSFPVVCQLHKSLGARPQ-----GDRMSRVVIE----- 175

145 KEFDESIRHTLVLSLQDAYKDRQIFRPLPLACHRL--KNAPEYVEVHSAENLFHSIYN 202

176 KPFGHDLASARELNKAVNAVPEEAVFR-----IDHYLGKETVQNILALRFANQLPDPINW 231

203 ---DDKOKKLFLLTNHRCITRLALTGGYKKGIGAAE--VRN-----LLATRPSSQL 249

232 AHYVDHVQ-----ITMAEDIGLGRAGYDYGIGAARDVIONHLMOLLALTAMEEPVSFH 285

250 DSYIMA-KVYVILASGAIIGNP-----QILYNSGFGS----- 278

286 PAALQAEKIKVLSATRLAEPLDQTSRGQYAGWGQGEKVGLLDEEGFAEDSTTETFAA 345

279 --LQVTPRNDLSIP-----NLGRYITEQPMAFCOIVLRQEFV--DSVRDDPYGLPMWK 327

346 ITLEVDTERRWAGVPFYLRTGKRLGRRVTEIALVF-----RRAPHLFPDATMTDELG----- 396

328 EAVAQHIAKNPFDALPIPFRRPEPQVTTPTTEEHFWHTQIHRDA---FSYGAV-----GPE 380

337 -----TNAMVIRVQ-PDEGVTLRFGSKVPGTAMEVRDVMNDFSYGSAFAEDSPE 444

381 VDSRVIVDL 389

445 AVERLLIDV 453

Search completed: October 28, 2003, 16:00:38

Job time : 48 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: October 28, 2003, 15:47:26 ; Search time 26 Seconds

(without alignments)
982.135 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618

Sequence: 1 NAEGRVAVPVGKHKNEI.....IINTLKSGTGDKNTGHEHRL 543

Scoring table: BIOSUM62

Gapop 10.0, Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1142.5	39.4	623	1	PROD CORVE
2	1119	38.6	622	1	PROD TRMT
3	1116	4.0	514	1	G6PD_MICTU
4	104.5	3.6	1337	1	PR2_DROME
5	101	3.5	1010	1	SCA4_RICPA
6	100	3.4	1142	1	ENAM_PIG
7	98	3.4	925	1	W70T_HUMAN
8	97.5	3.4	328	1	LPKX_YERPE
9	97.5	3.4	639	1	AMTG_HSPAK
10	97.5	3.4	1424	1	NCO3_HUMAN
11	97.5	3.4	1536	1	SIN3_YEAST
12	97	3.3	612	1	AMYG_ASPOR
13	96.5	3.3	902	1	MOPR_ARATH
14	96	3.3	291	1	USO2_HSV11
15	96	3.3	964	1	YINO_YEAST
16	95.5	3.3	519	1	AMTH_SACFI
17	95.5	3.3	708	1	HELS_SUTSO
18	95.5	3.3	1257	1	CCAA_BACTU
19	95	3.3	500	1	YDAK_YEAST
20	94.5	3.3	773	1	CDH_FHACH
21	94.5	3.3	855	1	GAP1_SCHPO
22	94.5	3.3	1054	1	R131_ARATH
23	93.5	3.2	406	1	YNOS_YEAST
24	93	3.2	470	1	LEU2_AZOVI
25	93	3.2	737	1	AMY1_ABDAR
26	93	3.2	890	1	GLND_ECOLI
27	93	3.2	890	1	GLND_SALTY
28	93	3.2	955	1	VP2_BTV17
29	93	3.2	1012	1	UBA1_SCHPO
30	92	3.2	376	1	DJBC_MOUSE
31	92	3.2	704	1	RPA1_BOVIN
32	92	3.2	813	1	KRAF_CABEL
33	92	3.2	890	1	GLND_SALTY

34	92	3.2	1584	1	U104_CABEL	P23678 caenorhabdi
35	92	3.2	1638	1	DPO3_LACIA	O9cdt7 lactococcus
36	91.5	3.2	425	1	IUCD_ECOLI	P13295 escherichia
37	91.5	3.2	840	1	MCMB_HUMAN	O9uf33 homo sapien
38	91	3.1	993	1	YAJ1_SCHPO	O09901 schizosach
39	90.5	3.1	847	1	VAV3_HUMAN	O09k24 homo sapien
40	90.5	3.1	1693	1	SAS_DROME	O04164 drosophila
41	90	3.1	463	1	GUNB_BACSP	P23019 bacillus sp
42	90	3.1	487	1	ENG4_CHLPN	O92762 chlamydia p
43	90	3.1	492	1	FTSA_HELPY	O25629 helicobacte
44	90	3.1	606	1	PRIM_MYXKA	P50070 myxococcus
45	90	3.1	737	1	PL02_MOUSE	O9f0b9 mus musculu

ALIGNMENTS

RESULT 1	PROD CORVE	STANDARD	PRT	623 AA.
AC	P79076			
DT	28-FEB-2003 (Rel. 41, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Pyranose oxidase precursor (EC 1.1.3.10) (PROD) (Glucose 2-oxidase).			
OS	Coriolus versicolor.			
OC	Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;			
OC	Apophalloporales; Coriolus.			
OX	NCBI_TaxID=57466;			
RN	[1]			
RP	SEQUENCE FROM N.A., SEQUENCE OF 38-63; 92-98; 186-207; 215-225;			
RC	378-411; 491-513 AND 611-621, CHARACTERIZATION, AND SUBUNIT.			
RA	STRAIN=Pe4a; TISSUE=Mycelium.			
RX	MEDLINE=97177816; PubMed=9025322;			
RT	Nishimura I., Okada K., Koyama Y.,			
RT	"Cloning and expression of pyranose oxidase cDNA from Coriolus			
RT	versicolor in Escherichia coli."			
RU	J. Biotechnol. 52:11-20(1996).			
CC	-1- FUNCTION: May play a role in lignin degradation of wood rot fungi			
CC	by supplying H(2)O(2) for the lignin peroxidases. Acts also on D-			
CC	xylose, L-sorbose, D-galactose and 1,5-anhydroglucitol.			
CC	-1- CATALYTIC ACTIVITY: D-glucose + O(2) = 2-dehydro-D-glucose +			
CC	H(2)O(2).			
CC	-1- COFACTOR: FAD; covalent.			
CC	-1- SUBUNIT: Homotrimer.			
CC	-1- SUBCELLULAR LOCATION: Hyphal periplasmic space (Potential).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: D73369; BAA1119.1.			
DR	InterPro: IPR00205; NAD_binding.			
KW	Oxidoreductase; Flavoprotein; FAD; Periplasmic.			
FT	PROPEP 37			
FT	CHAIN 38..623			
SQ	SEQUENCE 623 AA; 69495 MW; 5D3FC81B35FA5B54 CRC64;			
Query Match	39.4%; Score 1142.5; DB 1; Length 623;			
Best Local Similarity	45.1%; Pred. No. 1e-80;			
Matches 251; Conservative 69; Mismatches 163; Indels 73; Gaps 15;				
QY	13 GYHKKEIFKQIDRFVNVIGALQOVSVPVRNQNVPFLDGAASAPGSSAISNGKP 72			
DB	87 GAHKNTVYQKNIDFVNVIGALQOVSVPVNTLVIDPLSTSWQA--SSFFVRGSP 144			
QY	73 HOREPENTSAEAVTRGVGMSTHWTCTPRIHPPMESLPGIGRPKLSPNDPAE-DDKENNE 131			
DB	145 EDDPRLNLSGCAVTRVVGMSTHWTCTPRPRDQ-----RPLLVDDQDADDAEMDR 197			

QY 132 LYSEARLIGTSTKEFDESIRHTLVLRSLQDAYKDRQIRFPLPLACHRLKNAPEYVEWH 191
 DB 198 LYTKAESYFKTGTDQFQKESIRHNLVNLKLAEEYKG-QRDFQIPLAATR--RSPTFEVWS 254
 QY 192 SAENLP--HSYNDKOKKLLFTLLTNHRCRLALTGGYEKKIGAAEVRNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPDAPNERENLFPVACERV-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKVVLASGAINQOILVNSGSL-QVTPRN-DSLIPNLGRY 295
 DB 296 ESLIHDLISGRFEIKADVFLVLTAGAVHNAQLVNSGFGQLGRPDANPPQLPLPSLSY 355
 QY 296 ITEQPMAPCQIVLRQEFVDSVRDDPY--GLP-----WKEAQAQH 333
 DB 356 ITEQSLVFCQVMSDELIDSVKSDMIIRGNPDGLGYSVYTPGAETNKHDPWNEKVKNH 415
 QY 334 IAKNPTDALPIPRDPPEQVTTPTTEHPWHTQIHRDAFSYAGVPEVDSRVIVDLWFG 393
 DB 416 MNOHQEDPLPIPFEDPEPQVTTLFQSPHPWHTQIHRDAFSYAGVQVQSIDSLIYDWRFFG 475
 QY 394 ATDPEANLLVQNDVQGYSMPOPTFRYR--PSTASNVRAKMMADMCEVASNLGGYLPT 452
 DB 476 RTEPKENKLPFSDKITDYNMPOPTDFRPPAGRTSKAEADMTDMCMWSAKLGGFLPG 535
 QY 453 SPPQFMDPLGLHLAGTTRIGFKA--TTVADNNSLVMDFANLYVAGNGTIRTGFGENPT 510
 DB 536 SLPOFMEGLVHLGGTHRMGFDQEDKCCVNTDSRVFGKFLGCGGNIPATYAGNPT 595
 QY 511 LTSCHAIKARSIN 526
 DB 596 LTAMSLAIKSEYIKN 611

RESULT 2

PROD_TRAHI STANDARD; PRT; 622 AA.
 ID PROD_TRAHI
 AC P59097;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Pyranose oxidase precursor (EC 1.1.3.10) (PROD) (Glucose 2-oxidase).
 OS Trameetes hirsuta (White-rot fungus) (Coriolus hirsutus).
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Aphyllophorales; Trameetes.
 OX NCBI_TaxID=5327;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 38-62; 405-412; 486-491 AND 575-596.
 RC STRAIN=DSM 12119;
 RA Christensen S., Lassen S.F., Schneider P.;
 RT "Nucleic acids encoding polypeptides having pyranose oxidase activity.";
 RL Patent number US6146865, 05-MAY-1999.
 CC -!- FUNCTION: May play a role in lignin degradation of wood rot fungi by supplying H₂O(2) for the lignin peroxidases. Acts also on D-xylose, L-sorbose, D-mannose, D-galactose, 2-deoxy-D-glucose, methyl-alpha-D-glucopyranoside, methyl-beta-D-glucopyranoside, 5-keto-D-gluconic acid, and D-gluconic acid.
 CC -!- CATALYTIC ACTIVITY: D-glucose + O(2) = 2-dehydro-D-glucose + H(2)O(2).
 CC -!- COFACTOR: FAD; covalent (By similarity).
 CC -!- SUBCELLULAR LOCATION: Hypthal periplasmic space (Potential).
 DR InterPro: IPR000205; NAD binding
 KW Oxidoreductase; Flavoprotein; FAD; Periplasmic.
 FT PROPEP 1 37
 FT CHAIN 38 622 PYRANOSE OXIDASE.
 FT CONFLICT 60 60 T -> S (IN REF. 1; AA SEQUENCE).
 SQ SEQUENCE 622 AA; 69082 MW; 0B616C02014E7C5B CRC64;

Query Match 38.6%; Score 1119; DB 1; Length 622;
 Best Local Similarity 44.6%; Pred. NO. 6.8e-79;
 Matches 251; Conservative 69; Mismatches 183; Indels 60; Gaps 15;

QY 13 GYHKONEIEPQKIDRFPVNVIKGALQOVSVVPRNQNVTLPDGLAWSAPPOSSAISNGKP 72
 DB 87 GSHKNTVEYQKIDKIDFNVVQGLMEVSVVPMVMTVVDLTLSASWQA--STFFVRNGANP 144
 QY 73 HQREFENLSAEAVTRGVGGMSTHTWCTSPRIHPHMESLPGIGRPKL-SNDPAEDKKEWNE 131
 DB 145 EODPLRNLSCQAVTRVVGGMSTHTWCTATPFEK-----LORPLLVKNDKADDAEADR 197
 QY 132 LYSEARLIGTSTKEFDESIRHTLVLRSLQDAYKDRQIRFPLPLACHRLKNAPEYVEWH 191
 DB 198 LYTKAESYFKTGTDQFQKESIRHNLVNLKLAEEYKG-QRDFQIPLAATR--RSPTFEVWS 254
 QY 192 SAENLP--HSYNDKOKKLLFTLLTNHRCRLALTGGYEKKIGAAEVRNLLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPDAPNERENLFPVACERV-----VRN-----TSNSEI 295
 QY 250 DSY-----IMAKVVLASGAINQOILVNSGSL-QVTPRN-DSLIPNLGRY 295
 DB 296 ESLIHDLISGRFEIKADVFLVLTAGAVHNAQLVNSGFGQLGRPDANPPQLPLPSLSY 355
 QY 296 ITEQPMAPCQIVLRQEFVDSVRDDPY--GLP-----WKEAQAQH 333
 DB 356 ITEQSLVFCQVMSDELIDSVKSDMIIRGNPDGLGYSVYTPGAETNKHDPWNEKVKNH 415
 QY 334 IAKNPTDALPIPRDPPEQVTTPTTEHPWHTQIHRDAFSYAGVPEVDSRVIVDLWFG 393
 DB 416 MNOHQEDPLPIPFEDPEPQVTTLFQSPHPWHTQIHRDAFSYAGVQVQSIDSLIYDWRFFG 475
 QY 394 ATDPEANLLVQNDVQGYSMPOPTFRYR--PSTASNVRAKMMADMCEVASNLGGYLPT 452
 DB 476 RTEPKENKLPFSDKITDYNMPOPTDFRPPAGRTSKAEADMTDMCMWSAKLGGFLPG 535
 QY 453 SPPQFMDPLGLHLAGTTRIGFKA--TTVADNNSLVMDFANLYVAGNGTIRTGFGENPT 510
 DB 536 SLPOFMEGLVHLGGTHRMGFDQEDKCCVNTDSRVFGKFLGCGGNIPATYAGNPT 595
 QY 511 LTSCHAIKARSIN 526
 DB 596 LTAMSLAIKSEYIKN 611
 QY 402 LVPQNDVQGYSMPOPTFRYR--PSTASNVRAKMMADMCEVASNLGGYLPTSPQFMDPG 461
 DB 485 KLMFSDKITDAYNLQPTDFR--FPGREAEADMTDMCMWSAKLGGFLPGSPQFMEPG 542
 QY 462 LALHLAGTTRIGFKA--TTVADNNSLVMDFANLYVAGNGTIRTGFGENPTLTSCHAIK 519
 DB 543 LVHLGGTHRMGFDQEDKCCVNTDSRVFGKFLGCGGNIPATYAGNPTLTSCHAIK 602
 QY 520 SARSINLTKGGTGDKNVTGEHRN 542
 DB 603 SCEVI---KCNFSPNPVKHNN 622
 RESULT 3
 G6PD MYCTU STANDARD; PRT; 514 AA.
 ID G6PD MYCTU
 AC O08407;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Glucose-6-phosphate 1-dehydrogenase (EC 1.1.1.49) (G6PD).
 GN ZWF OR ZWF2 OR RV1447C OR MT1494 OR MTCV493.07.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RA Cole S.T., Broesch R., Parkhill J., Garnier T., Churcher C., Harris D., Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F., Badcock K., Basham D., Brown D., Chillingworth T., Connor R., Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S., Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L., Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J., Rutter S., Seeger K., Skelton S., Squares S., Squares R., Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the complete genome sequence.";
 RT Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;

Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 Peterson J., DeBoy R., Dodson R., Gwin M.L., Haft D., Hickey E.,
 Kolonay J.F., Nelson W.C., Mayman L.A., Emolaeva M.D., Salzberg S.L.,
 Delcher A., Uterback T., Weidman J., Khouri H., Gill J., Mikulga A.,
 Bishai W.;
 "Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains";
 Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate + NADP(+) = D-glucono-
 1,5-lactone 6-phosphate + NADPH.
 -1- PATHWAY: Pentose phosphate pathway; first step.
 -1- SIMILARITY: BELONGS TO THE GLUCOSE-6-PHOSPHATE DEHYDROGENASE
 FAMILY.
 -1- CAUTION: M.TUBERCULOSIS HAS TWO GENES FOR ZWF. THIS ONE LOOKS LIKE
 A CLASSICAL ZWF.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL: Z95844; CAB09259.1; -
 EMBL: AE007019; RAK45757.1; -
 PIR: B70917; B70917.
 HSSP: P11411; 1DPG.
 TIGR: MT1494; -
 TubercuList; RV1447C; -
 InterPro: IPR001282; G6PD.
 Pfam: PF00479; G6PD_C.1.
 Pfam: PF02781; G6PD_C.1.
 PRINTS: PR00079; G6PDHGNASE.
 ProDom: PD001129; G6PD; 1.
 TIGRFAMs: TIGR00871; zwf; 1.
 PROSITE: PS00069; G6P DEHYDROGENASE; 1.
 Oxidoreductase; NADP; Glucose metabolism. Complete proteome.
 ACT_SITE 210 210 BY SIMILARITY.
 SEQUENCE 514 AA; 57343 MW; 22A9CCDC9AB062F CRC64;

 Query Match 4.0%; Score 116; DB 1; Length 514;
 Best Local Similarity 22.8%; Pred. No. 0.26;
 Matches 98; Conservative 48; Mismatches 163; Indels 120; Gaps 23;
 y 29 FVNVKALQV-VSVVRNQNVTLPDGAWSAPPGSAISNGKPNPHOREFENLSAEAVTR 87
 b 77 FGQVYNVAQEHCTPPFQQNWDRLAEG-FRVPFGTFDDDDAFQAETLEKLDAAE---R 132
 y 88 GVGGMSTHTWCSTPRIHPPM-ESL--PGIGRPKLSNDPAEDDKENNELYSAEERLIGTST 144
 b 133 GTGGNHAFYLAIPKSPFVVCQELHKSGLARPQ-----GDRWSRVVIE----- 175
 y 145 KEFDESIRHTLVLSLDAYKDRQIRPLPLACHRL--KNAPEVEVHNSAENLPHSTYN 202
 b 176 KPGHDLASARELNKAVNAVPEEBAVFR-----IDHLYGKVTQNLARFANFQDFDPIWN 231
 y 203 ----DDQKKLFTLLNHRCTRLALTGYEKKIGAAE--VFN-----LLATRNPSQL 249
 b '232 AHVYDVHQ-----ITWAEIDIGLGRAGYDYGIGARDVQIHLMLQALLTAMEEVSFH 285
 y 250 DSYTMA-KVYVLASGATGNP-----QILYNSGFSG----- 278
 b 286 PAALQAEKIKVLSATRLAELPDQTTSRQYAAQWQGEKVVGLLDEGFAEDSTTETFAA 345
 y 279 --LQVTRNDSLIP-----NLGRYTEQPMACQIVLRQEFV--DSVRDDPVLGLPWK 327
 b 346 ITLEVDFRWRAGVFFYLRGKRLGRVTEIALVF-----RRAPHLFPDATMTDELG---- 396
 y 328 EAVAQIAKTNPTDALPFPFRDPEPQVTPPTFEEHPHTQIHRDA---FSYGAV---GPE 380
 b 397 -----TNAMVIRVQ-PDEGVTLRFSGKVPFGTAMEVRDVMNMFDSYGSFAEDSPSE 444

QY 381 VDSRVTVDL 389
 DB 445 AYERLILDV 453

 RESULT 4
 PR2_DROME STANDARD; PRT; 1337 AA.
 AC Q917F7; Q24316; Q96026; Q9V6K0;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Tyrosine-protein kinase PR2 (EC 2.7.1.112).
 GN PR2 OR CG3969.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID:7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkelley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazew R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 Lasoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 Murtulov G., Mishaev N.V., Mobarry C., Morris J., Moshrefi A.,
 Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
 Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
 Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
 Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 RN [2]
 REVISIONS, AND ALTERNATIVE SPLICING.
 RC STRAIN=Berkelley;
 RX MEDLINE=22426063; PubMed=12537572;
 RA Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
 Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
 Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
 Bettencourt B.R., Celisner S.E. de Grey A.D.N.J., Drysdale R.A.,
 Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
 Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
 Lewis S.E.;
 RT "Annotation of the Drosophila melanogaster euchromatic genome: a

systematic review." ;
 Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
 [3]
 RP SEQUENCE FROM N.A. (ISOFORM LONG).
 RC STRAIN=Berkeley; TISSUE=Embryo;
 RA Stapleton M., Brokstein P., Hong L., Agbavani A., Carlson J.W.,
 RA Champe M., Chavez C., Dorsett V., Farfan D., Friese E., George R.A.,
 RA Gonzalez M., Guarin H., Li P.W., Liao G., Miranda A., Mungall C.J.,
 RA Nuno J., Pacleb J.M., Paragas V., Park S., Phouanavong S.,
 RA Wan K.H., Yu C., Lewis S.E., Rubin G.M., Celiniker S.E.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE OF 13-917 FROM N.A. (ISOFORM LONG).
 RP STRAIN=Oregon-R; TISSUE=Embryo;
 RC MEDLINE=94156202; PubMed=8118607;
 RX Ito M., Mateui T., Taniguchi T., Chihara K.;
 RA "Alternative splicing generates two distinct transcripts for the
 RT Drosophila melanogaster fibroblast growth factor receptor homolog." ;
 RL Gene 139:215-218(1994).
 [5]
 RN SEQUENCE OF 266-321 FROM N.A.
 RP MEDLINE=98401146; PubMed=9731193;
 RA Oates A.C., Wollberg P., Achen M.G., Wilks A.F.;
 RT "Sampling the genomic pool of protein tyrosine kinase genes using the
 RT polymerase chain reaction with genomic DNA." ;
 RL Biochem. Biophys. Res. Commun. 249:660-667(1998).
 CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
 CC tyrosine phosphate.
 CC -|- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=Long;
 CC IsoId=Q917F7-1; Sequence=Displayed;
 CC Name=Short;
 CC IsoId=Q917F7-2; Sequence=VSP_005006;
 CC -|- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES.
 CC -|- SIMILARITY: Contains 1 CRIB domain.
 CC -|- CAUTION: Ref. 4 sequence differs from that shown due to frameshifts
 CC in positions 752, 806, 816 and 827.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AE003819; AAF58423.4; -;
 DR EMBL; AE003819; AAC22275.3; -;
 DR EMBL; AY051755; AAX93179.1; -; FRAME.
 DR EMBL; D17551; BAA04489.1; ALT_FRAME.
 DR EMBL; AJ002909; CAA05744.1; -;
 DR HSSP; P11362; IFGK.
 DR FlyBase; FBgn0013955; P2.
 DR GO; GO:0004713; F:protein tyrosine kinase activity; NAS.
 DR GO; GO:0006468; P:protein amino acid phosphorylation; NAS.
 DR InterPro; IPR000095; FAKbox/Rho-binding.
 DR InterPro; IPR000719; Prot_kinase.
 DR InterPro; IPR001245; Tyr_kinase.
 DR Pfam; PF00018; SH3; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR ProDom; PD000001; Prot_kinase; 1.
 DR SMART; SM00285; PBD; 1.
 DR SMART; SM00220; S_TKC; 1.
 DR SMART; SM00219; TyRK; 1.
 DR PROSITE; PS00108; CRIB; 1.
 DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
 DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
 KW Transferase; Kinase; Tyrosine-protein kinase; ATP-binding;
 KW Alternative splicing.
 FT DOMAIN 133 399
 FT NP_BIND 139 147
 FT ATP (BY SIMILARITY).

FT BINDING 164 164 ATP (BY SIMILARITY).
 FT ACT_SITE 260 260 BY SIMILARITY.
 FT DOMAIN 488 502 CRIB.
 FT VARSPLIC 1 82 Missing (in isoform Short).
 FT CONFLICT 450 450 /FTid=VSP_005006.
 FT CONFLICT 484 484 F -> FSTFWKGLSTGKTGYF (IN REF. 4).
 FT CONFLICT 906 906 R -> A (IN REF. 4).
 FT CONFLICT 913 913 A -> P (IN REF. 4).
 FT CONFLICT 913 913 D -> V (IN REF. 4).
 SQ SEQUENCE 1337 AA; 147475 MW; 12650E9DF5A82724 CRC64;
 Query Match 3.6%; Score 104.5; DB 1; Length 1337;
 Best Local Similarity 20.7%; Pred. No. 8.5;
 Matches 132; Conservative 77; Mismatches 228; Indels 201; Gaps 34;
 QY 5 GTAPVYVPGYHKQNEIEFQKIDRFVNVVVKALQOVSVVVRNQNVP-----TLDPGAMS 58
 DB 48 GLSRPEIRLRKRYEKHFPHS---YLSKIKRLQLQAPCTVMVKREAPGGGQVALD-GSSA 103
 QY 59 APGSSAISNG-----KNPHQREFENLSAEAVTRGVG---GMSTH--WTCSTPRIHPPM 107
 DB 104 SACSSLAANKGASSPSKVPNNKHIIIPADSLSVNKQLGTGEGFIVQGVWNGNERIOVAI 163
 QY 108 ESLPGIGRKLNDPAEDDKENNELYS-EAERLI-----GTSTKEFDESIRHTLVLSR 159
 DB 164 KCL---CRERMOSNPWEFLKEAAMHSIEHENIVRLYGVLATDSLMLVTELAH---LRS 217
 QY 160 LDQAYKDR--QRIFRPLPLACH---RLKNAPEYVEVHSAENLFHS-----IYNDDQ 206
 DB 218 LLECLKDGLRVSLFTIPTLCFALQICGMRYLE---QKRLIHRDLAARNILVFSKDKV 274
 QY 207 KKLFTLLTHRCTR-LALTGGYEK-----KIG-----AAEVNLLATRNPSOLDYIM 254
 DB 275 K-----ISDFGLSRALGVGKDYKTNFNVNKLPIAWCAPECINYLRFTHASD----- 322
 QY 255 AKVYVLAGSAGNQPQLYN---SGFSGLOVTPRNDLSIPNLGRYITEQP----- 300
 DB 323 ---VWAFGVCLWEMPSYGFQFWAALTGLQILEIDA--PNYOR--LEQDDCCPSEYVTL 374
 QY 301 MAPC-----QIVLRQEFVDSVRDDPYG 322
 DB 375 MNKQWDDAAKPRFGEIYDQPMKPEQLKAVVNVCTEPKKDHLRYQGDIISVLDRTNG 434
 QY 323 LPWKEAVAQHTAK--NPT-----DALPIPRDPPEOVTPPTTEHPHWTOI-----HRD 370
 DB 435 TPFWKGVLSTGTGYGTPNPSTVAFLEGLPSSTRDSFVSVDHRSRKRKLETEMISKPDND 494
 QY 371 AFSYGAVGPEVDSRVIVDLRWFGAT-----DPEANNLLVFQNDVQDG 412
 DB 495 FKHTGHVG--IDGATFGDIAFLGSSQNYNHVPKQIVTPKPSEDI EOTPLLL----- 544
 QY 413 YSMPQPTFRYRPSTASNVNRKRMADMCEVASNLGGYLPTSPFPQMDPGIALHLAGTTRI 472
 DB 545 --PPTPTSPDSLQTAGSYFPE-----GANSNGAMGTS-----KNPTFIPSAEHTPKL 589
 QY 473 GFDKATTVADNNSLVWDFP-----NLTVAGNGTIRTCFG 506
 DB 590 -----IATNGQSSPFDFAAGSTNPFNRRGDDLEFG 620
 RESULT 5
 SC44_RICPA
 ID SC44_RICPA STANDARD; PRT; 1010 AA.
 AC Q9AJ75;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Antigenic heat-stable 120 kDa protein (PS120) (120 kDa antigen)
 DE (protein P120) (fragment).
 GN SC44 OR D.
 OS Rickettsia parkeri.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rickettsiales;
 OC Rickettsiaceae; Rickettsiaseae; Rickettsia.

NCBI_TaxID=35792;
 [1]
 SEQUENCE FROM N.A.
 Sekeyova Z., Roux V., Raoult D.,
 "Phylogenetic analysis of Rickettsia spp. by comparing sequence of the
 'gene D' coding for an intracytoplasmic protein.",
 Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; AF155059; AAK30690.1; -
 Antigen.
 1
 NON_TER 1010 1010
 T T
 Q SEQUENCE 1010 AA; 110694 MW; 7858608C92FF9C5B CRC64;
 Query Match 3.5%; Score 101; DB 1; Length 1010;
 Best Local Similarity 18.2%; Pred. No. 10;
 Matches 106; Conservative 88; Mismatches 243; Indels 144; Gaps 25;
 61 PGSSAISNGKNPHREFENLSAEAVTRGVGMSTHWCTSPR-----IHPMESLPQIGRP 116
 55 FMSVLSGNISPSQ-----TSDPITKAV-----RETIQPKQNLIEQILKDLAAL--- 100
 117 KLSNDPAEDDKENWLYSEAEELIGT-----STKFP-DESIHRTLVLSLQD-----AYKD 166
 101 -TDRDLAQKQKEIEEKEKDTLSTFFGNPANREFIDKALENPELKKLESIEIAGYKN 159
 167 RORIPRPLPLACHRLKNAPEYVEWH---SAENLFHSIYNDKQKLFTH----- 212
 160 VHNTPS-----AASGVPGGFKPQWENHVSASDLRATVKNDAAGSELCTNETVTVKPFT 215
 213 LTNHRCRLALTGGYE-----KKGAEEVRNLLATR-----NPSOLDYSYIMAKVYVL 260
 216 LAKQDGTQVOISSYREIDFPKLDKADGSMHL-SNVALKADGTSKDKPVYFTAH---Y 271
 261 ASGAINPQILYNSGFGLOVTPRNDSLIPNLGRVITPQPAFCQIVLRQEFVDSV----- 316
 272 EEGPQKPKLEISSPKLKFAGTGDDAI-----AVIEHGGEIYTLAVTRGKYKDMWKEVE 327
 317 -----RDDPYGLPWKEAVQAQIAKNPTDALPIPRDPPEQVT 354
 328 LNOGSVDLSQAEIIICGOSKOEPLTP---OQTSSVEPPQHKQVPPITPTNQPL 384
 355 TPFTTEHPWHQIHH-----RDAPSYGAVG--PEVDSRVTVDLRWFGAT--- 395
 385 OPEISQPSQVNPVNLNTALSGSQMDLLVYNVAGLTVEIDSNKQIDLIKEAATAIL 444
 396 -----DPEANNLVFQNDVQDGYSNPQPTFR-----YAPSTASNVRAKQNMAD 438
 445 NNEKSDIVEKQANTIAAENTVNNKLNKLPDAKVAGVNAVLEIINKDQNTNPLEKSKMLEA 504
 439 MCEVASNLGGVLPSTPPQFM-----DPGLALHAGTTRIGFKATTVADNNSLVWDFANLY 494
 505 TVAILVNSLENLEPKQKQKQMLEKAVDVGLSL-----KDDASRAATIDIGIKOVVIK-SNLY 557
 495 VAGNGTTRTGFGENTPLTSMCHAIKSARSIIINTLKGDTGK 535
 558 TEDKGTMLIAVGDKVNVSELSNAEKQ-KLIGSVLKKGVQEAQ 597

 RESULT 6
 _ENAM_PIG
 STANDARD; PRT; 1142 AA.
 O97939;
 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Enamelin precursor.
 GN ENAM.
 OS Sus scrofa (Pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 OC NCBI_TaxID=9823;
 RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RC TISSUE=Enamel epithelium;
 RX MEDLINE=98040070; PubMed=9372788;
 RA Hu C.-C., Fukae M., Uchida T., Qian Q., Zhang C.H., Ryu O.H.,
 RA Tanabe T., Yanakoshi Y., Murakami C., Dohi N., Shimizu M.,
 RA Simmer J.P.;
 RT "Cloning and characterization of porcine enamel mRNA";
 RL J. Dent. Res. 76:1720-1729(1997).
 RN [2]
 RP SEQUENCE OF 39-773 FROM N.A., SEQUENCE OF 39-49; 174-276;
 RP 515-524; 535-578; 641-646; 663-665; 670-686; 740-750; 765-773 AND
 RP 833-848, AND CHARACTERIZATION.
 RX MEDLINE=97350624; PubMed=9206327;
 RA Fukae M., Tanabe T., Murakami D., Dohi N., Uchida T., Shimizu M.;
 RT "Primary structure of porcine 89' kDa enamel";
 RL Adv. Dent. Res. 10:111-118(1996).
 CC -!- FUNCTION: Involved in the mineralization and structural
 CC organization of enamel. Involved in the extension of enamel during
 CC the secretory stage of dental enamel formation.
 CC -!- SUBCELLULAR LOCATION: Secreted. Extracellular matrix.
 CC -!- TISSUE SPECIFICITY: EXPRESSED BY SECRETORY-PHASE AMELOBLASTS.
 CC INTACT ENAMELIN AND LARGE-MOLECULAR-WEIGHT ENAMELINS ARE LIMITED
 CC TO THE MOST SUPERFICIAL LAYER OF THE DEVELOPING ENAMEL MATRIX,
 CC WHILE LOW-MOLECULAR-WEIGHT ENAMELINS ARE OBSERVED IN DEEPER
 CC ENAMELIN. PREFERENTIAL LOCALIZATION AMONG THE CRYSTALLITES IN ROD
 CC AND INTEROD ENAMEL.
 CC -!- DEVELOPMENTAL STAGE: EXPRESSED FROM LATE DIFFERENTIATION TO THE
 CC TRANSITION STAGE.
 CC -!- PTM: PARENT ENAMELIN IS PROTEOLYTICALLY CLEAVED INTO SEVERAL
 CC SMALLER POLYPEPTIDES. CLEAVAGE OF N-TERMINAL REGION OF ENAMELIN
 CC OCCURS SOON AFTER SECRETION.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U52196; AAD10837.1; -
 PIR; T37455; T37455.
 GlycoSuiteDB; O97939;
 GO; GO:0005578; C:extracellular matrix; ISS.
 GO; GO:0030345; F:structural constituent of tooth enamel; ISS.
 GO; GO:0030282; P:bone mineralization; ISS.
 GO; GO:0042476; P:odontogenesis; ISS.
 Biomineralization; Extracellular matrix; Glycoprotein; Signal;
 Hydroxylation; Phosphorylation.
 SIGNAL 1 38
 FT CHAIN 39 1142 ENAMELIN.
 FT CHAIN 56 KDA ENAMELIN.
 FT CHAIN 89 KDA ENAMELIN.
 FT CHAIN 142 KDA ENAMELIN.
 FT CHAIN 155 KDA ENAMELIN.
 FT CHAIN 174 276 32 KDA ENAMELIN.
 FT CHAIN 515 665 25 KDA ENAMELIN.
 FT CHAIN 670 34 KDA ENAMELIN.
 FT CHAIN 45 KDA ENAMELIN.
 FT MOD_RES 53 53 PHOSPHORYLATION (PROBABLE).
 FT MOD_RES 191 191 PHOSPHORYLATION.
 FT MOD_RES 216 216 PHOSPHORYLATION.
 FT MOD_RES 547 547 HYDROXYLATION.
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .).

FT CARBOHYD 252 252 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 264 264 N-LINKED (GLCNAC. . .)
 FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 462 462 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 929 929 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CARBOHYD 1040 1040 N-LINKED (GLCNAC. . .) (POTENTIAL)
 FT CONFLICT 680 680 H -> D (IN REF. 2)
 FT CONFLICT 838 840 RDH -> TTI (IN REF. 2)
 SQ SEQUENCE 1142 AA; 128352 MW; 938306BC87CC3FC6 CRC64;

Query Match 3.4%; Score 100; DB 1; Length 1142;
 Best Local Similarity 19.2%; Pred. No. 15;
 Matches 118; Conservative 65; Mismatches 210; Indels 220; Gaps 32;

QY 1 NAEGETAVPYV-GYH-----KNEIEPKQIDR-----FNVVKG 35
 DB 192 NEEGGN--PYRGFFGYHGGRPPYSEEMFEQDFEPKPKDPKTPATEPSVN---245
 QY 36 ALQOVSVVRNVPYL-DP-GAWSAPPSSAISGNKNPHQR-----EFENLSAEAVTR 87
 DB 246 ----TTVPETNSTQPNAPNPGNDTSPTGTS--CGGNPRSNPTGQNGPAVNVSGQGVPR 299
 QY 88 ----GVGGMSTHWTCTSPRI-----HPMESLFGIGR-----PKLSNDPAEDDKW 129
 DB 300 SQSPWGPRTQITIHENYPENINRPFARRQWPPG---PAMGHRNRPFFYRNQIQORPRW 356
 QY 130 NELYSEABRLIGTSTKFDSEIRITLVLSLODAYKORIFRPLPLACHRLKNAPYVE 189
 DB 357 NSFTLEGG-----QAVRPGYPTIRYVY-----GSTARSNPPNYA- 390
 QY 190 WSAENLFHSIYNDKQKLLFTLTJNHRCTRLALTG-----GYEKKIGAAE 235
 DB 391 -GNSANLRKKEGPNK-----NPMVTNVAIPPGRKHGTVDQENIQNPREKQVSQKE 440
 QY 236 VRNLLATRNPS-----SOLDSYIMAKVYVLASGAIGNPQILYNSGFGSLQVTPRNDSLIPN 291
 DB 441 -RTVVPTRDPGPNRSODYIGKSNYKL-----PQ-----PEDNMLVFN 479
 QY 292 LG-----RYTEOPMAFCQIVLQKQFVDSVRDPPYGLPWKEA---VAQ 332
 DB 480 FNSIDORENSYYPRESKRAFNDSGQTQTIIPK-----GIVLEPRPIPYESETNOPELK 534
 QY 333 HIAKNP--TDALPIPFRRDEP-----QVTFPTEHPHWTQIHRDAFSGYGVAGPEV 381
 DB 535 HSAVQPVVTEGIPSPAKEHPPAGNTWNQOEISPPFKED--PGRQEEHLPHLSHG-----587
 QY 382 DSRVIVDLRFMGATDPEANLLVFQNDVQDGSMP-----OPTFRRPSTASNVRAKQMA 437
 DB 588 -SRHVYYPDNPYPDPRENSPYLSNTWYERDDSPNTMGQDENPHYPMWTPDPKET-----642
 QY 438 DMCEVASNLGYLTPSPQFMDPGLALHLACTTTRIGFDKATTADVADNNSLVWDFANLYVAG 497
 DB 643 -----IPYNEEDPIDTGDHFPQQR-----WDMBEELSFK 674
 QY 498 NGTIRTGGENPT 510
 DB 675 DPTVRHYEGEQYT 687

RESULT 7
 W70T_HUMAN

ID W70T_HUMAN STANDARD; PRT; 925 AA.
 AC P57737;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE 70 kDa WD-repeat tumor rejection antigen homolog.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.

RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
 Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
 Tanaka T., Nakamura Y., Isogai T., Sugano S.;
 RA "NEDO human cDNA sequencing project.";
 RT Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RL -! SIMILARITY: Contains 8 WD repeats.
 CC -! SIMILARITY: BELONGS TO THE CORONIN FAMILY OF WD-REPEAT PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AK025674; BAB15211.1; -
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF00400; WD40; 6.
 DR PRINTS; PR00320; GPROTEINRPT.
 DR SMART; SM00320; WD40; 7.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 DR PROSITE; PS0082; WD_REPEATS_2; 4.
 DR PROSITE; PS0294; WD_REPEATS_REGION; 2.
 KW Repeat; WD repeat.
 FT REPEAT 75 115 WD 1.
 FT REPEAT 124 163 WD 2.
 FT REPEAT 166 205 WD 3.
 FT REPEAT 209 253 WD 4.
 FT REPEAT 542 582 WD 5.
 FT REPEAT 592 632 WD 6.
 FT REPEAT 635 674 WD 7.
 FT REPEAT 728 768 WD 8.
 SQ SEQUENCE 925 AA; 100574 MW; A393CE973C94F4A9 CRC64;

Query Match 3.4%; Score 98; DB 1; Length 925;
 Best Local Similarity 20.2%; Pred. No. 16;
 Matches 88; Conservative 52; Mismatches 139; Indels 156; Gaps 21;

QY 54 PGAWSAPPSSAISGNKNPHQREFENLSAEAVTRGV-----GGMSTHWTCTSPRIHPME 108
 DB 108 PGGQALPSAGPVVLG--PEDLPVEVLQFHPTSDGLVSAAGTTVKVMDAAKQO---PLT 162
 QY 109 SLPGICRPLNSNDPAEDDKENNELYSEAEELICTSTKEFDESIRHTLVLSLODAYKDRQ 168
 DB 163 EL-----AAHGDVLVQSAVSRDGLVGT-----ACKDQK 191
 QY 169 -RIFRPLPLACHRLK-NAPEYVEWH-----SAENLFHSIYNDKQKLL-----209
 DB 192 LRIFDP-----RTKPRASQSTOAHENSRSRLAWMGTWELVSTGNGREREVKLWD 245
 QY 210 -----FTLTNHRCT-----RLALTGGYEKKIGAAEVRNLLATRNPSQ--L 249
 DB 246 RPFSSALASITLDTSLGCLVPLDPSGLLVLAGKGRQLCYEVVFPQALSPVTCVL 305
 QY 250 DSYIMAKVYVLASGAIGNPQILYNSGFGSLQVTPRNDSLIPNLGRYTEQPMACQIVLR 309
 DB 306 ES-----VLRGAALVPRQALAVMGCEVLRLVQLSDTAIVPIGYHVRKAVEFHEDL--356
 QY 310 QEFVDSV---RDDPYGLPW---KEAVAQHIKNP-----TDALPIPFRRDPEQVTT 355
 DB 357 --FPDTAGCVATDPDTHG--WWAGDNQOVQKVSLNPACRPHPSFTSCL-VPPAEPLPTAQ 411
 QY 356 PFTEHPHWTQIHRDAFSGYGVAGPEVDSRVIVDLRFMGATDPEANLLVFQNDVQDGSMP 415
 DB 412 PAVMETP-----VGDADASEGFS 430
 QY 416 PQTFRRPSTASNV 430
 DB 431 P-PSLTSPTSPSSL 444

RESULT 8

56 AWSAP-----PGSSAISNGKNP-----HQREFENLSAEAVTRGVGMSTHW----- 96

DR PRINTS: PR00736; GLHYDELASE15.
 DR PRODOM: PD001568; CBD_4; 1.
 DR POSITE; PS00820; GLUCOAMYLASE; 1.
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW SIGNAL; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 24
 FT CHAIN 25 639
 FT DOMAIN 494 538
 FT BINDING 143 143
 FT ACT SITE 199 199
 FT ACT_SITE 202 202
 FT ACT_SITE 203 203
 FT ACT_SITE 233 236
 FT DISULFID 245 472
 FT DISULFID 285 293
 FT CARBOHYD 194 194
 FT CARBOHYD 418 418
 FT CARBOHYD 464 464
 FT CARBOHYD 466 466
 FT CARBOHYD 467 467
 FT CARBOHYD 475 475
 FT CARBOHYD 476 476
 FT CARBOHYD 482 482
 FT CARBOHYD 483 483
 FT CARBOHYD 485 485
 FT CARBOHYD 487 487
 FT CARBOHYD 491 491
 FT CARBOHYD 495 495
 FT CARBOHYD 498 498
 FT CARBOHYD 499 499
 FT CARBOHYD 500 500
 FT CARBOHYD 501 501
 FT CARBOHYD 503 503
 FT CARBOHYD 505 505
 FT CARBOHYD 507 507
 FT CARBOHYD 511 511
 FT CARBOHYD 512 512
 FT CARBOHYD 513 513
 FT CARBOHYD 514 514
 FT CARBOHYD 516 516
 FT CARBOHYD 517 517
 FT CARBOHYD 519 519
 FT CARBOHYD 521 521
 FT CARBOHYD 523 523
 FT CARBOHYD 524 524
 FT CARBOHYD 525 525
 FT CARBOHYD 526 526
 FT CARBOHYD 527 527
 FT CARBOHYD 528 528
 FT CARBOHYD 529 529
 FT CARBOHYD 530 530
 FT CARBOHYD 531 531
 FT CARBOHYD 533 533
 FT CARBOHYD 534 534
 SQ SEQUENCE 639 AA; 68271 MW; E112B31A4D8D06B CRC64;
 Query Match 3.4%; Score 97.5; DB 1; Length 639;
 Best Local Similarity 20.2%; Pred. No. 9.9;
 Matches 106; Conservative 51; Mismatches 194; Indels 175; Gaps 25;
 Qy 80 LSAEAVTRGVGGMSTHWTCTPPIHPMPESLPGLGRPKLSNDPAEDDKENNELYSARL 139
 Db 106 ISSQAITQGVNPSGDL-----SSGGLGEPKFNVDYATVGTSGMRPQRDPAL 153
 Qy 140 IGTSTKEFDESIRHTLVLSIQDAYKD-RQRIERPLPLACHRLKNAPEYVE--WHAENL 196
 Db 154 RATAMIGFGQVL-----LDNGYTSAAATEIWNPL-----VRNDSLYVAQYWNQGT-- 197
 Qy 197 FHSIYNDKOKKFLTLTHNRCTRALTGG--YEKIGIAAEVRNLLATRNPPSSQLDSYIM 254
 Db 198 -YDLMEWVNGSSPFTIAVQHR-----ALVEGSATATVAGSS-----CSWCDSSQA- 240

QY 255 AKVYVLASGAIGNPOIL--YNSGFSGLQVTPRNDSLPN-----LGRVITEQPMAPCO- 305
 Db 241 -----PQLCLVQFWTGSYILANPDSRRSGKDTNLTLSGTHTFDFEACDD 287
 QY 306 -----IVLRQEFVDSR-----DDPY--GLPWKCAVAQ 332
 Db 288 STFPQCSPRALANHKVEVDSPRSIVTLNDGLSDSEAVAVGRYPEDSYNGNPNFQSTLA- 346
 QY 333 HIAKMTDALPIPRDPPEQVTPPTTEHPWHTQ-----IHRDAFSYGAVG--- 378
 Db 347 -AAEQLYDAL-----YQWQKQSGLEITDVSDFKALYSGAATGYS 387
 QY 379 --PEVDSRVIVDLRWFG-----ATDPBANLLVFQNDVQDG--XSMPOPTFYRPSSTA 427
 Db 388 SSSSYSSIVSAVKTFADGFVSIETHAASNSLSQDFKSDGDELSARDLTSYAAALLT 447
 QY 428 SNVRARKMM-----ADMCEVAGNLCGY-----LPTSPQFMDPGLALHLAGTTRI 472
 Db 448 ANNRNRNVVPPSWGETSASSWPGTCAATSASGYSSVTVTWSPSIVATGGTTTTATTGTS 507
 QY 473 G----FDKATTVADNNLSLWDFAN-----LYVAGNGTIRTGGEN 508
 Db 508 GGVTSKTTTTSKTTTTSSTTSCTTPTTAVAVTFLDTATTYGEN 553
 RESULT 10
 NC03 HUMAN
 ID NC03 HUMAN STANDARD; PRT; 1424 AA.
 AC Q9Y6Q9; Q9BR49; Q9UPC9; Q9UPG4; Q9UPG7;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Nuclear receptor coactivator 3 (EC 2.3.1.48) (NCoA-3) (Thyroid hormone
 DE receptor activator molecule 1) (TRAM-1) (ACTR) (Receptor-associated
 DE coactivator 3) (RAC-3) (Amplified in breast cancer-1 protein) (AIB-1)
 DE (Steroid receptor coactivator protein 3) (SRC-3) (CBP-interacting
 DE protein) (PCIP)
 GN NC03 OR TRAM1 OR RAC3 OR AIB1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND INTERACTION WITH CREBBP;
 RP PCAF; RARA; RXRA; THRA AND ESR.
 RX TISSUE=Pituitary;
 RX MEDLINE=98010595; PubMed=9346901;
 RA Takeehita A., Cardona G.R., Koibuchi N., Suen C.-S., Chin W.W.;
 RT "TRAM-1, a novel 160-kDa thyroid hormone receptor activator molecule,
 RT exhibits distinct properties from steroid receptor coactivator-1.";
 RL J. Biol. Chem. 272:27629-27634(1997).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4), ENZYMATIC ACTIVITY, AND VARIANT
 RP 1248-GLN-GLN-1250 DEL.
 RC TISSUE=Leukemia;
 RX MEDLINE=97410321; PubMed=9257036;
 RA Chen H., Lin X.-J., Schiltz R.L., Chakravarti D., Nash A., Nagy L.,
 RA Privalsky M.L., Nakatani Y., Evans R.M.;
 RT "Nuclear receptor coactivator ACTR is a novel histone
 RT acetyltransferase and forms a multimeric activation complex with p/CAF
 RT and CBP/p300.";
 RL Cell 90:569-580(1997).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM 5), AND INTERACTION WITH ESR.
 RC TISSUE=Lung;
 RX MEDLINE=97400625; PubMed=9252329;
 RA Anzick S.L., Kononen J., Walker R.L., Azorsa D.O., Tanner M.M.,
 RA Guan X.-Y., Sauter G., Kallioniemi O.-P., Trent J.M., Meltzer P.S.;
 RT "AIB1, a steroid receptor coactivator amplified in breast and ovarian
 RT cancer.";
 RL Science 277:965-968(1997).
 RN [4]
 RP SEQUENCE FROM N.A. (ISOFORM 5), INTERACTION WITH VDR; RARA; PPARA;

RXRA; THRA AND ESR, AND VARIANT 1248-GLN--GLN-1250 DEL.
 TISSUE=Brain;
 MEDLINE=97385128; PubMed=9238002;
 Li H., Gomes P.J., Chen J.D.;
 "RAC3, a steroid/nuclear receptor-associated coactivator that is
 related to SRC-1 and TIF2.";
 Proc. Natl. Acad. Sci. U.S.A. 94:8479-8484(1997).
 [5]
 SEQUENCE FROM N.A.
 MEDLINE=21638749; PubMed=11780052;
 Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
 Jones M., Scavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
 Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,
 Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
 Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
 Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
 Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
 Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,
 Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
 Graffham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,
 Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
 Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
 Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
 Lehaeaslaio M.H., Leverisha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
 Marsh V.L., Martin S.L., McConachie L.J., McIay K., McMurray A.A.,
 Milne S.A., Mistry D., Moore M.J.P., Mullikin J.C., Nickerson T.,
 Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
 Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
 Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Showkeen R., Sims S.,
 Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
 Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
 Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
 Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A.,
 Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
 Rogers J.;
 "The DNA sequence and comparative analysis of human chromosome 20.";
 Nature 414:865-871(2001).
 [6]
 ACETYLATION BY CREBBP, AND MUTAGENESIS OF LYS-616; 619-LYS--LYS-620;
 LYS-647; LYS-681; LYS-687; LYS-700 AND LYS-708.
 MEDLINE=99418638; PubMed=10490106;
 Chen H., Lin R.J., Xie W., Wilpitz D., Evans R.M.;
 "Regulation of hormone-induced histone hyperacetylation and gene
 activation via acetylation of an acetylase.";
 Cell 98:675-686(1999).
 [7]
 INTERACTION WITH NFKB1.
 MEDLINE=20547534; PubMed=11094166;
 Werbach S., Nojek I., Lanz R., Costas M.A.;
 "RAC-3 is a NF-kappa B coactivator.";
 PERS Lett. 485:195-199(2000).
 [8]
 SUBUNIT OF A COMPLEX CONTAINING CREBBP, NCOA2; IKKA; IKKB AND IKKBG,
 AND PHOSPHORYLATION.
 MEDLINE=21968797; PubMed=11971985;
 Wu R.-C., Qin J., Hashimoto Y., Wong J., Xu J., Tsai S.-Y., Tsai M.-J.,
 O'Walley B.W.;
 "Regulation of SRC-3 (pCIP/ACTR/AIB-1/RAC-3/TRAM-1) coactivator
 activity by I kappa B kinase.";
 Mol. Cell. Biol. 22:3549-3561(2002).
 [9]
 VARIANTS IN POLY-GLN REGION.
 MEDLINE=98394734; PubMed=9727751;
 Shirazi S.K., Bober M.A., Coetzee G.A.;
 "Polymorphic exonic CAG microsatellites in the gene amplified in
 breast cancer (AIB1 gene)."
 Clin. Genet. 54:102-103(1998).
 -1- FUNCTION: Nuclear receptor coactivator that directly binds nuclear
 receptors and stimulates the transcriptional activities in a
 hormone-dependent fashion. Plays a central role in creating a
 multi-subunit coactivator complex, which probably acts via
 remodeling of chromatin. Involved in the coactivation of different
 nuclear receptors, such as for steroids (GR and ER), retinoids

(RARs and RXRs), thyroid hormone (TRs), vitamin D3 (VDR) and
 prostanoids (PPARs). Displays histone acetyltransferase activity.
 Also involved in the coactivation of the NF-kappa-B pathway via
 its interaction with the NFKB1 subunit.
 -1- CATALYTIC ACTIVITY: Acetyl-CoA + histone = CoA + acetyl-histone.
 -1- ENZYME REGULATION: Coactivator activity on nuclear receptors and
 NF-kappa-B pathways is enhanced by various hormones, and the TNF
 cytokine, respectively. TNF stimulation probably enhances
 phosphorylation, which in turn activates coactivator function. In
 contrast, acetylation by CREBBP apparently suppresses coactivation
 of target genes by disrupting its association with nuclear
 receptors.
 -1- SUBUNIT: Present in a complex containing NCOA2, IKKA, IKKB, IKKBG
 and the histone acetyltransferase protein CREBBP. Interacts with
 PCAF.
 -1- SUBCELLULAR LOCATION: Mainly cytoplasmic and weakly nuclear. Upon
 TNF activation and subsequent phosphorylation, it translocates
 from the cytoplasm to the nucleus.
 -1- ALTERNATIVE PRODUCTS:
 Event=Alternative splicing; Named isoforms=5;
 Comment=Additional isoforms seem to exist;
 Name=1;
 IsoId=Q9Y6Q9-1; Sequence=Displayed;
 Name=2;
 IsoId=Q9Y6Q9-2; Sequence=VSP_003407;
 Name=3;
 IsoId=Q9Y6Q9-3; Sequence=VSP_003405, VSP_003407, VSP_003408;
 Name=4;
 IsoId=Q9Y6Q9-4; Sequence=VSP_003405, VSP_003406, VSP_003407,
 VSP_003408;
 Name=5;
 IsoId=Q9Y6Q9-5; Sequence=VSP_003408;
 -1- TISSUE SPECIFICITY: Widely expressed. High expression in heart,
 skeletal muscle, pancreas and placenta. Low expression in brain,
 and very low in lung, liver and kidney.
 -1- DOMAIN: Contains three leu-xaa-xaa-leu (LXXLL) motifs. Motifs
 1 and 2 are essential for the association with nuclear receptors,
 and constitute the RID domain (Receptor-interacting domain).
 -1- PTM: Phosphorylated by IKK complex and acetylated by CREBBP.
 Acetylation occurs in the RID domain, and disrupts the interaction
 with nuclear receptors. Both modifications regulate its function.
 -1- POLYMORPHISM: The length of the poly-Gln region is polymorphic in
 the normal population.
 -1- MISCELLANEOUS: NCOA3 is frequently amplified or overexpressed in
 breast and ovarian cancers.
 -1- SIMILARITY: BELONGS TO THE SRC/P160 FAMILY OF NUCLEAR RECEPTOR
 COACTIVATORS.
 -1- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
 -1- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
 -1- CAUTION: Ref.5 (CAC36067) sequence differs from that shown due to
 erroneous gene model prediction.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
 or send an email to license@isb-sib.ch.

 EMBL; AF016031; AAC51849.1; -
 EMBL; AF036892; AAB92368.1; -
 EMBL; AF012108; AAC51677.1; -
 EMBL; AF010227; AAC51663.1; -
 EMBL; AL034418; CAC17693.1; -
 EMBL; AL034418; CAC36067.1; ALT_SEQ.
 PIR; T03851; T03851.
 PDB; 1KBH; 06-FEB-02.
 Genew; HGNC:7670; NCOA3.
 MIN; 601937; -
 GO; GO:0005634; C:nucleus; NAS.
 GO; GO:0016923; F:ligand-dependent thyroid hormone receptor i. . .; NAS.
 GO; GO:0003713; F:transcription co-activator activity; NAS.

DR GO:0006350; P:transcription; NAS.
 DR InterPro; IPR001092; HLH_Basic.
 DR InterPro; IPR000014; PAS_domain.
 DR Pfam; PF00010; HLH; 1.
 DR Pfam; PF00989; PAS; 1.
 DR SMART; SM00353; HLH; 1.
 DR SMART; SM00091; PAS; 1.
 DR PROSITE; PS00038; HLH_1; FALSE_NEG.

Query Match 3.4%; Score 97.5; DB 1; Length 1424;
 Best Local Similarity 19.1%; Pred. No. 33;
 Matches 102; Conservative 64; Mismatches 210; Indels 159; Gaps 23;

QY 41 SVVNRNVPLDP-----GAWAPGSSAISKPH-QREFN-----LSAEA 84
 DB 960 TLPLRSNIPGARVLOQQOQMLQWRGEIPMGANGPYQAAASNLGWSWPDGMLSMQ 1019
 QY 85 VTRGVGMSHTWCTSTRIHPMESLPGIGRPKLSNDPAEDDKENNELYSEAEELIGTST 144
 DB 1020 V-----SHGTQNRPLRNSLDL--VGPPSNEGSDERALLDQLHTLLSNWDATGL 1069
 QY 145 KEFD-----ESIRHTLVLSQDAYKDRQRIFR-----PLPLACHRL 181
 DB 1070 EIDRALGIPLVNQGALEPKQAFQGEAAVMDQKAGLYGQTYPAQGPWGGFHLQ 1129
 QY 182 KNAPEYVHSAENLFHSIYNDKQKLFLLTNHRTLRALTCGYEKKIQAABVRNLLA 241
 DB 1130 GQSPS-----FNSMMNQMGQNFPLQGMH-----PRANIMR 1161
 QY 242 TR-NPSSQSDSYIMAKYVVLASGAINQPILYNSGSLQVTPNDLSLIPNLGYITEOP 300
 DB 1162 PRTPKQLRMQLQRL-----QCGQFLNQSQAELKLMEN-----PTAGGAAMRP 1208
 QY 301 MAFQCIIVLRQEFVDSVRDDPYGLPWLKAEVAQHLAKNPDALPIPR----- 347
 DB 1209 MMQFQVSSQGFN-----AQWVQRSELLSHHFQQRVAMWMMQOQQQ 1252
 QY 348 ---DPRFQVTPFTEHPMTQHHRDAFSGVAGVDSRVIVDLRWFGATDPEANNLLV 404
 DB 1253 QQQ 1307
 QY 405 FQNDVQGYSM---PQTFRVRPTASVNRKVMADNCEVASNLGSLYLTSPQFWDPG 461
 DB 1308 -QFPYQPNYGMGQQQDFAFG-RVSSPPNAM-----MSSRMG---FSQNPMMQHPQ 1352
 QY 462 LALHLAGTTRIGFDKATTVADNNSLVMDFANLYVAGNGTIRTGFGENTILTSMSCH 516
 DB 1353 AASIYQSSEMKGWFSG-NLARNSS-----FSQQQFAHQG-----NPAYYSWVH 1394

RESULT 11
 SIN3 YEAST
 ID SIN3 YEAST STANDARD; PRT; 1536 AA.
 AC P22579; Q08049;
 DT 01-AUG-1991 (rel. 19, Created)
 DT 30-MAY-2000 (rel. 35, Last sequence update)
 DT 30-MAY-2000 (rel. 39, Last annotation update)
 DE Paired amphipathic helix protein.
 GN SIN3 OR SD11 OR UME4 OR RPD1 OR CAM2 OR SDS16 OR YOL004W.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=GRF88.
 RA MEDLINE=91042523; PubMed=2233725;
 RT Wang H., Clark I., Nicholson P.R., Herskowitz I., Stillman D.J.;
 RT "The Saccharomyces cerevisiae SIN3 gene, a negative regulator of HO,
 RT contains four paired amphipathic helix motifs.";
 RL Mol. Cell. Biol. 10:5927-5936 (1990).
 RN [2]
 RP SEQUENCE FROM N.A.

RA Hughes B., Pohl T.M.;
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: REQUIRED FOR BOTH FULL TRANSCRIPTION REPRESSION AND
 CC ACTIVATION OF MANY GENES INCLUDING CELL TYPE-SPECIFIC GENES (STE6,
 CC TY2 AND HO), CELL DIFFERENTIATION-SPECIFIC GENES (SPO13), GENES
 CC THAT RESPOND TO EXTERNAL SIGNALS (PHO5) AND TRK2. IT IS PROBABLY
 CC INVOLVED IN THE SAME TRANSCRIPTIONAL REGULATORY FUNCTION OR
 CC PATHWAY AS THE TRANSCRIPTIONAL REGULATORY PROTEIN RPD3
 CC -1- SUBUNIT: IT PROBABLY FORMS A COMPLEX WITH THE TRANSCRIPTIONAL
 CC REGULATORY PROTEIN RPD3.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR, POSSIBLY LINKED TO CENTROMERE.
 CC -1- DOMAIN: CONTAINS 4 PAIRED AMPHIPATHIC HELIX MOTIFS, SEPARATED BY A
 CC 10 TO 30 AA SEGMENT THAT FORMS POSSIBLY A LOOP, RESULTING IN A
 CC STRUCTURE SIMILAR TO THAT OF HLH AND TPR MOTIFS.
 CC -1- SIMILARITY: TO S.POMBE SPAC12C2.10C.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).

DR EMBL; M36822; AAA34839.1; -;
 DR EMBL; Z74746; CAA99003.1; -;
 DR PIR; S66686; RGSYS3.
 DR TRNSFPAC; T03597; -;
 DR SGD; S0005364; SIN3.
 DR GO; GO:000118; C:histone deacetylase complex; IPI.
 DR GO; GO:0004407; F:histone deacetylase activity; IPI.
 DR GO; GO:0006347; P:chromatin silencing at RML and HMR (sensu S.); IMP.
 DR GO; GO:000183; P:chromatin silencing at ribosomal DNA (rDNA); IMP.
 DR GO; GO:0006348; P:chromatin silencing at telomere; IMP.
 DR GO; GO:0006357; P:regulation of transcription from Pol II pro.; IGI.
 DR InterPro; IPR003822; PAH.
 DR Pfam; PF02671; PAH; 3.
 DR Repeat; Transcription regulation; Repressor; Cell division;
 KW Activator; Nuclear protein.
 FT DOMAIN 239 1200 4 X 2 PAIRED (A,B) AMPHIPATHIC HELICES.
 FT REPEAT 239 252 HELIX A.
 FT REPEAT 274 285 HELIX B.
 FT REPEAT 426 439 HELIX A.
 FT REPEAT 461 472 HELIX B.
 FT REPEAT 679 692 HELIX A.
 FT REPEAT 714 725 HELIX B.
 FT REPEAT 1152 1165 HELIX A.
 FT REPEAT 1189 1200 HELIX B.
 FT DOMAIN 480 519 GLN-RICH.
 FT CONFLICT 510 510 O -> QAQ (IN REF. 2).
 SQ SEQUENCE 1536 AA; 174838 MW; 0834726312B13878 CRC64;

Query Match 3.4%; Score 97.5; DB 1; Length 1536;
 Best Local Similarity 19.4%; Pred. No. 37;
 Matches 103; Conservative 71; Mismatches 163; Indels 193; Gaps 28;

QY 58 SAPFGSSAISNGKPHQREFNLSA-----EAVTRGVGMSHTWCTSTRIHPMESLP 111
 DB 1044 SGSDGSSIASRQAPYQEMSLDLHRSRYOKLR-----SNDECKVPQSPPEEP 1098
 QY 112 G-IGRPKLSNDPAEDD-----KENNELYSEA 136
 DB 1099 NTIEBELIDEAKNPLTGNLVEANSOGIIQNRISFNLFANTNIYIFPRHWTIY--- 1155
 QY 137 ERLIGTSTKEFDESIRHTLVLSL-----QDAYKQRIRFP 173
 DB 1156 ERLT--EIKOMNERTVKEINTRVTVTFAKDILLSSQLSEMGDFVEDAYKQVILSR- 1212
 QY 174 LPLACHRLKNAPEYVHSAENLFHSIYNDKQKLFLL-----LTNHRCTRLALTCGY 227
 DB 1213 -----RLINGDLEHQWF--EESLRQAYN-NKAFKLYTIDKVTQSLVKHAHT--LMTDAK 1261
 QY 228 EKKIGAAEVRNLATRNPPSSQLDSYIMAKYVVLASGAINQPILYNSGSLQVTPNDLS 287

1262 TABIMAFVXD-----RN-----ASTTSKQIIYR-----LQVR-SHMS 1295
 288 LIPNLGR-----YITEQPMACQIVLRQEFVDSVRDDPYGLPWKVAQHIANKPT 339
 1296 NTEMFRIEFDKTLHVSIOYIALDDTLKEPKADEK-----WKYYVTSYALPHPT 1347
 340 DALPPIPRDPEQVTPFTTEHPWHTQIHRDAFSYGAVGPEVDSRVIVDLRWFGATDPEA 399
 1348 EGI-----PHEKKIPFLB-----RLIEGQDIDGTEVEDEP-----SPEG 1383
 400 NNL-----LVFQNDVODG-YSM--POPTRYRSTASNVRRARMMADMCEVASNLG 447
 1384 ISVTLKIKIQTIPYQIHLIENGSDYVETKATNKY-PTIANDNTQKMGWSQKELIS--- 1439
 448 GYLPTSPPPFMD--PGLALHLAGTTRIGDK-----ATTVAONNSL 486
 1440 -----KFLDCAVGLRNNLDEAQLSMOKWENLKDSIAKTSAGNOGI 1481

RESULT 12

MYG ASPOR
 ID MYG ASPOR STANDARD; PRT; 612 AA.
 AC P36914;
 IT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 JT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Glucoamylase precursor (EC 3.2.1.3) (Glucan 1,4-alpha-glucosidase)
 BE (1,4-alpha-D-glucan glucohydrolase).
 GLAA.
 ASpergillus oryzae.
 NC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
 SC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
 XC NCBI_Taxid=5062;
 UN [1]
 IN SEQUENCE FROM N.A.
 PX MEDLINE=91254744; PubMed=1368680;
 P Hata Y., Kitamoto K., Gomi K., Kumagai C., Tamura G., Hara S.;
 A "The glucoamylase cDNA from Aspergillus oryzae: its cloning,
 T nucleotide sequence, and expression in Saccharomyces cerevisiae.";
 L Agric. Biol. Chem. 55:941-949(1991).
 L [2]
 P SEQUENCE FROM N.A.
 P MEDLINE=92104497; PubMed=1761224;
 A Hata Y., Tsuchiya K., Kitamoto K., Gomi K., Kumagai C., Tamura G.,
 X Hara S.;
 A "Nucleotide sequence and expression of the glucoamylase-encoding gene
 (glaa) from Aspergillus oryzae.";
 T Gene 108:145-150(1991).
 L [3]
 P SEQUENCE FROM N.A.
 P STRAIN=RIB 40;
 A Hata Y., Tamura G., Kumagai C., Gomi K., Kitamoto K., Tsuchiya K.,
 A Hata Y.;
 L Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
 C -1- CATALYTIC ACTIVITY: Hydrolysis of terminal 1,4-linked alpha-D-
 C Glucose residues successively from non-reducing ends of the chains
 C with release of beta-D-glucose.
 C -1- SIMILARITY: BELONGS TO FAMILY 15 OF GLYCOSYL HYDROLASES.
 C
 C This SWISS-PROT entry is copyright. It is produced through a collaboration
 C between the Swiss Institute of Bioinformatics and the EMBL outstation -
 C the European Bioinformatics Institute. There are no restrictions on its
 C use by non-profit institutions as long as its content is in no way
 C modified and this statement is not removed. Usage by and for commercial
 C entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 C or send an email to license@isb-sib.ch).
 C
 C EMBL; D01035; BAA00841.1;
 C EMBL; D10698; BAA01540.1;
 C PIR; J01346; J01346.
 C HSP; P04064; ICAI
 C InterPro; IPR002044; CBD_4.

DR InterPro; IPR000165; Glyco_hydro_15.
 DR Pfam; PF00686; CBM_20; 1.
 DR Pfam; PF00723; Glyco_hydro_15; 1.
 DR PRINTS; PR00736; GLHYDRLASE15.
 DR PRODOM; PD001568; CBD_4; 1.
 DR PROSITE; PS00820; GLUCOAMYLASE; 1.
 KW Hydrolase; Glycosidase; Polysaccharide degradation; Glycoprotein;
 KW Signal.
 FT SIGNAL 1 19 BY SIMILARITY.
 FT PROPEP 20 25 BY SIMILARITY.
 FT CHAIN 26 612 GLUCOAMYLASE.
 FT BINDING 146 146 SUBSTRATE (BY SIMILARITY).
 FT ACT_SITE 202 202 CATALYTIC BASE (BY SIMILARITY).
 FT ACT_SITE 205 205 GENERAL ACID CATALYST (BY SIMILARITY).
 FT ACT_SITE 206 206 INTERACT WITH SUBSTRATES (BY SIMILARITY).
 FT DISULFID 236 239 BY SIMILARITY.
 FT DISULFID 248 475 BY SIMILARITY.
 FT CARBOHYD 39 39 N-LINKED (GLUCNA...) (POTENTIAL).
 SQ SEQUENCE 612 AA; 65486 MW; CD7B23E5FA978F97 CRC64;
 Query Match 3.3%; Score 97; DB 1; Length 612;
 Best Local Similarity 20.5%; Pred. No. 10;
 Matches 105; Conservative 56; Mismatches 161; Indels 190; Gaps 28;
 QY 78 ENLSAEAVTRGV-----GGMSTHTWCSTPRIHPMESLPGIGRPKLSNDPAEDDKENNELY 133
 DB 107 EFISQARIQISNPSSGALSS-----GGLGEPKENVDETATGAWGRPQ 150
 QY 134 SAEERLIGTSTKEFDESI---RHTVLRLSDAYKDRORIFRPLPLACHRLKNAPYEVW 190
 DB 151 RDGPALRATAMISFGEWLVSHTSIATDL-----VMPV-----VRNLSYVAQ 194
 QY 191 HSAENLFHSIYNDKQKLLFTLLTNHCTRLLATGG--YEKKIG-----AAEVRNL 239
 DB 195 YWSQSGP-DLMEVQCTSFVAVSHR---ALVEGSSFAKTVGSSCPYCDQAQVRCY 249
 QY 240 LATRNPPSQLDSTYIMAKYVVLASGAIGNPQILYNSGFGSLQVTPRNDLSIPNLGYITEQ 299
 DB 250 L-----QSFWTGSIQAN-----FGGGRSGKQIN-----TVLGSIHFTD 283
 QY 300 PMAFQIVLRQEFVDSVRDDPYGLPWKVAQHIANKPTDALPIPP----- 346
 DB 284 FOATC-----DDATFQPCSAALANH--KVVTDSFRSIYAINSGRAENQAVA 328
 QY 347 --RDEPQVTTPTTEHPWH-----TQIHRDAFSYGAVGPEVDSRVIVD--LRFW--- 392
 DB 329 VGRYPEDS-----YNGNPWFLLTTLAAEQLDALYQWDKIG-----SLAITDVSLEPFFKAL 380
 QY 393 ---GATDPEANNNLLVFQNDVQ-----DGYSMPQPTFRYRSTASNVRRARMMADMCEVA 443
 DB 381 YSSAATGTYSSTTVYKDIVSAVKAYADGYVQIVQT--YAASTGS-----MAEQ----- 427
 QY 444 SNLGGYLPSPQPMQDGLALHLAGTTRIGDKATTVADNNNSLVNDVFANLYVAG---NGT 500
 DB 428 ---YTKTDGSO-----TSARD---LTWSYAALLTANNRRNAV 458
 QY 501 IRTGFGEN-----ATLASMCHAIKSARSINI 527
 DB 459 VPAPWGETAATSIPTASCTTSASGTYSVIT 490

RESULT 13

MOPT ARATH
 ID MOPT ARATH STANDARD; PRT; 902 AA.
 AC P93024; O64965; Q9FXI3;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Transcription factor MONOTEROS (Auxin-responsive protein IAA24)
 DE (Auxin response factor 5).
 GN MP OR IAA24 OR ARF5 OR ATIG19850 OR F6F9_8 OR F6F9.10.
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 [1]
 RN SEQUENCE FROM N.A., AND VARIANT ILE-796.
 RP STRAIN=cv. Columbia;
 RC STRAIN=cv. Columbia; PubMed=9482737;
 RX MEDLINE=98151364; PubMed=9482737;
 RA Hardtke C.S., Berlet T.;
 RT "The Arabidopsis gene MONOPTEROS encodes a transcription factor
 mediating embryo axis formation and vascular development.";
 RL EMBO J. 17:1405-1411(1998).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC STRAIN=cv. Columbia; PubMed=11130712;
 RX MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
 White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
 Buehler E., Chan A., Chao Q., Chen H., Cheuk R.P., Chin C.W.,
 Chung M.K., Conn L., Conway A.B., Creasy T.H., Dewar K.,
 Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin B.,
 Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 Militscher J., Miranda M., Nguyen M., Niemman W.C., Osborne B.I.,
 Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:816-820(2000).
 [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=cv. Columbia;
 RC STRAIN=cv. Columbia;
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 Goldsmith A.D., Lee J.M., Toziumi M.J., Yu G., Brooks S.Y., Chao Q.,
 Chen H., Karlin-Neumann G., Kim C.J., Lam B., Miranda M., Nguyen M.,
 Palm C.J., Shinn P., Southwick A.M., Davis R.W., Ecker J.R.,
 Theologis A.;
 RT "Full Length cDNA of gene F69.10 (GI:10086486).";
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
 [4]
 RN SEQUENCE OF 13-902 FROM N.A.
 RP MEDLINE=98004476; PubMed=9342315;
 RX Kim J., Harter K., Theologis A.;
 RT "Protein-protein interactions among the Aux/IAA proteins.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:11786-11791(1997).
 CC -1- FUNCTION: Transcriptional regulator that modulates gene activities
 in response to auxin signals. Involved in vascular strand
 formation.
 CC -1- SUBUNIT: Capable of forming a dimer with IAA1. Interacts with
 IAA17/AXR3.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- TISSUE SPECIFICITY: Embryo axis, provascular tissues, procambium
 and some differentiated vascular regions of mature organs.
 CC -1- DEVELOPMENTAL STAGE: In early embryo and during organ development.
 CC -1- MISCELLANEOUS: Absence of the protein causes early embryonic
 lethality. Premature stop codons are associated with vascular
 defects.
 CC -1- SIMILARITY: Contains 1 TF-B3 domain.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE AUX/IAA
 FAMILY.
 CC -1- CAUTION: Ref.2 sequence differs from that shown due to erroneous
 gene model prediction.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AF037228; AAC39410.1; -;
 DR EMBL; AF037229; AAC60794.1; -;
 DR EMBL; AC007797; AAG12546.1; ALT_SEQ.
 DR EMBL; AF334716; AAG50094.1; -;
 DR EMBL; U79557; AAB92476.1; -;
 DR PIR; T51807; T51807.
 DR TRANSFAC; T04507; -;
 DR InterPro; IPR003311; AUX_1AA.
 DR InterPro; IPR003340; TF_B3.
 DR Pfam; PF02309; AUX_1AA; 1.
 DR Pfam; PF02362; B3; 1.
 DR KW Multigene family; Nuclear protein; Transcription regulation;
 KW DNA-binding.
 FT DOMAIN 119 306 TF-B3
 FT DOMAIN 722 902 AUX/IAA-LIKE
 FT DOMAIN 23 28 POLY-THR.
 FT DOMAIN 282 289 POLY-ALA.
 FT DOMAIN 445 499 MET-RICH.
 FT VARIANT 796 796 T -> I.
 SQ SEQUENCE 902 AA; 99650 MW; D20D348369757F09 CRC64;
 Query Match 3.3%; Score 96.5; DB 1; Length 902;
 Best Local Similarity 18.8%; Pred. No. 20;
 Matches 83; Conservative 63; Mismatches 133; Indels 163; Gaps 21;
 QY 14 YHKKEIEFEQKIDIRFVNIKG-----ALQGVSVPRNQNVPTLDPGA--- 56
 DB 459 HNNQNVPSFSEMQQ--NIVMGNGLLGDMKMQOPLMMKQKSEWQPNKLTVPNSASNT 515
 QY 57 -----WSAP--PGSSAI---SNGKNPHQREFENLSAEAVTRGVGGMSTHTCSTP 101
 DB 516 SGQEQNLQSMSAPAKPENSTLSGSSGRVQHGLESMEQASQVT-----ISTVCNEE 568
 QY 102 RHPPMESLPDGIGRPKLSNDPAEDDK---EWNELYS-EAEELIGTSTKEFDEIRHTLVL 157
 DB 569 KVNQLQK-PG-----ASSPVQADQCLDITHIQYQPSDPIGFSFLEDTLSQVSSP 621
 QY 158 RSLQDAYK-----DRQRIFRPLACHRLKNAPYEVVHSAENLFHSIYND----- 204
 DB 622 QSLAGSYKOPFLLSSQDSSAVVLP-----DSTNSP-----LFHDVWDITQLNGLK 665
 QY 205 -----KQKLFLLTNRCTRLALTGGYKIGAAVRNLLATRNPSQLDSYIMAKV 257
 DB 666 FDFQFSLMOODLYA--SQNICMSNTTS-----NILDPLSLNTVLDDFCAIK- 710
 QY 258 YVLASGAIGNPQILYNSGFGSLQVTPRNDLIPNLGRYTEQPMACQIVLRQEFVD--- 314
 DB 711 -----DTDFQNHPSGGLVGNNTSFADQVQSQITSASFADSQAFSRQDFPDNSG 759
 QY 315 -----SVRDDPYGLPWWKEAQAQIAKNPDTALPIPRDPSPQVTTPTTEE 360
 DB 760 GTGTSSNVDFDCLSRQNSGSSWOKIAT-----PRVKT----- 794
 QY 361 HPWHTQIHRDAFSGYGAVGPEVD 382
 DB 795 ---YTKVQKT-----GSVGRSID 809
 RESULT 14
 US02 HSV11
 ID US02 HSV11 STANDARD; PRT; 291 AA.
 AC P06485;
 DT 01-JAN-1988 (Rel. 06, Created)
 DT 01-JAN-1988 (Rel. 06, Last sequence update)
 DT 01-NOV-1995 (Rel. 32, Last annotation update)
 DE Protein US2.
 GN US2.
 OS Herpes simplex virus (type 1 / strain 17).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;

OC Alphaherpesvirinae; Simplexvirus.
 RX NCBI_TaxID=10299;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=85160822; PubMed=2984429;
 RA McGeoch D.J., Doan A., Donald S., Rixon F.J.;
 RA "Sequence determination and genetic content of the short unique
 RT region in the genome of herpes simplex virus type 1";
 J. Mol. Biol. 181:1-13(1995).
 RL [1]
 CC -1- SIMILARITY: BELONGS TO THE HERPESVIRUS US2 FAMILY.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; L00036; AAA9686.1; --
 DR EMBL; X14112; CAA32279.1; --
 DR EMBL; X02138; CAA26056.1; --
 DR PIR; A05238; Q08E72.
 DR InterPro; IPR003485; US2_Unk.
 DR Pfam; PF02476; US2.1.
 DR SFQ SEQUENCE 291 AA; 32470 MW; 7227D9F051FD3A78 CRC64;

Query Match 3.3%; Score 96; DB 1; Length 291;
 Best Local Similarity 21.0%; Pred. No. 4;
 Matches 73; Conservative 31; Mismatches 113; Indels 130; Gaps 17;

QY 236 VRNLLATRN--PSQLD-----SYIMAKVYVLASGAIGNPQILYNSGSLQ----- 280
 Db 8 VMTLLDNALPRTSDASPALMSFLRQRIASPLGPGVVRPANRLRLAEPLMDLP 67

QY 281 -----VTPRNSLPIR--LGRYITQPMFAFCOIVLRQEFVDSV-----RDPD----- 320

Db 68 KPTPRIVTRSCRPNTITGLPAESDLE-----STEVDAVACFRLHRDQSPPL 121

QY 321 YGLPWWKEA-----CHIAKNPTD-----ALPFPDPPEQVTP 356

Db 122 YHL--WVVGADLCVPFEVYQAQIRLGVRFAIAKTPDAWGEWAVPTRLPEWTVAWTP 179

QY 357 F--TEEPHWHQIHRDAFSGVAGPEVDSRVIVDLRFQATDEANNLLVFQNDVQDGS 414

Db 180 FPAAPHPLLETLRSYEQYGVVLPCTNGRGRDCRWL-----RSLIALHK----- 225

QY 415 MPQPTFRYPSTASNVRAKRMWADMCEVASNLGGYLPSTP---PQMPDPLGLHLAGTTR 471

Db 226 -PHPA--TPCPLTSHVPRPCA--CMGPEVDPDEQPTSGRGPGQTDPLIAR----- 275

QY 472 IGFDKATTVDNNSLVDPANLYVAGNGTIRTFGNGENPTLISMCHAI 518

Db 276 -----GERPLPHICYPV 288

RESULT 15

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

1NO YEAST

RX

RA

RA

RA

RA

RA

RT

RL

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

PubMed=9169870;
 Churcher C.M., Bowman S., Badcock K., Bankier A., Brown D.,
 Chillingworth T., Connor R., Devlin K., Gentles S., Hamlin N.,
 Harris D.E., Hensell T., Hunt S., Jagels K., Jones M., Lye G.,
 Moule S., Odell C., Pearson D., Rajandream M.A., Rice P., Rowley N.,
 Skelton J., Smith V., Walsh S., Whitehead S., Barrell B.G.;
 "The nucleotide sequence of Saccharomyces cerevisiae chromosome IX";
 Nature 387:84-87(1997).
 RL [1]
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- SIMILARITY: Contains 1 Zn(2)-Cys(6) fungal-type binuclear cluster
 CC domain.
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; Z38059; CAA86148.1; --
 DR PIR; S48404; S48404.
 DR HSSP; P08657; 1CLD.
 DR SGD; S0001392; YII130W.
 DR InterPro; IPR001139; Fungi_TrN.
 DR Pfam; PF04082; Fungal_trans; 1.
 DR Pfam; PF00172; Zn clus; 1.
 DR SMART; SM00066; GAL4; 1.
 DR PROSITE; PS00463; ZN2_Cy6_FUNGAL_1; 1.
 DR PROSITE; PS00448; ZN2_Cy6_FUNGAL_2; 1.
 DR KW Hypothetical protein; Transcription regulation; DNA-binding;
 KW Nuclear protein; Zinc; Metal-binding.
 FT DNA_BIND 21 47 ZN(2)-CYS(6); FUNGAL-TYPE.
 FT DOMAIN 811 896 ASN-RICH.
 SQ SEQUENCE 964 AA; 108780 MW; AD5ABE59E4B022CC CRC64;

Query Match 3.3%; Score 96; DB 1; Length 964;
 Best Local Similarity 19.5%; Pred. No. 24;
 Matches 100; Conservative 76; Mismatches 163; Indels 174; Gaps 27;

QY 13 GYHKK-----NEIFQK-----DIDFVNVIKALQOVVPVRNQNVPTLD----- 53

Db 380 GPHKRLSPNSGFSPIEMRKRLFTYIKLDYINAMLGPRISIPDDPDTLP-LDLS 438

QY 54 -----PGAWSAPPGSSAISNGKNPHQREF-----ENLSAEAVTRGVGGMSTHWTCT 100

Db 439 ENITEVAVLPENQHSVLSSTGIS---NEHTKFLILNELIISLYPIKTSNIISHTVTS 495

QY 101 --PRIHPMESLPGIGRPKLSN-DPAEDDKEMNELYSEARLIGTSTKEFDESIRHTLV 157

Db 496 LELKRLNWLDSLPKELIPNAENIDPE-----YERANRLHL-----SFLHVQII 539

QY 158 RSLQDAYKDRQRIFRPL-----PLACHRLKN-----APEYVE----- 189

Db 540 -----LYRPFTHYLSRNNAENVDPLCYRRARNSTAVARTVIKAKEMVSNL 587

QY 190 ---:WHAENLFSIY-----NDKQKCLFTLTNRHCTRLALTGGYEKIG 232

Db 588 LTGSYHYACIYIFSVAGLLFYIHEAQLPKDSAREYDILKDAETGRSVLIQLKSSNA 647

QY 233 AAEVRNLLATRNPSQLDSYIMAKVYVLASGAIGNPQILY--SGFSLQVTPRNDSLIP 290

Db 648 ASRTVNL-----NQIFELKNSKIQLTA-----LHSSPSNESAFVLTNNSSALKP 693

QY 291 NLGRYITEQPMFAFCOIVLRQEF-----VDSVRDPYGLPWWKEAQAQIAKNTDLP 345

Db 694 HLGDSL-QPPVFFSSQDTKNSFLAKSEESTND-----YAMANYLNNTPISENPLN 743

QY 346 FRDPEPV---TTFTEHPHWHQIHRDAFSGVAGPEVDSRVIVDLRFQATDEANN 402

Db 744 EAQODQVSQGTNNMSNE-----RDPNNFLSIDIRLDNN-----GOSNI 782

QY 403 LVFQNDV---QDGYSMFPQPT---FRYRPSSTASN 429

Db 783 LDATDDVFIRNDG-DIPTNSAFDFSSKSNASN 814

Search completed: October 28, 2003, 15:57:59
Job time : 32 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.
OM protein - protein search, using sw model
Run on: October 28, 2003, 15:47:46 ; Search time 103 Seconds
(without alignments)
1360.413 Million cell updates/sec

Title: US-09-856-327-2_COPY_76_618
Perfect score: 2901
Sequence: 1 NAEETAVPVVPGYKKNEL.....IINTLGGTDGKNTGHRNL 543

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL 23:*
1: sp arChaea.*
2: sp bacteria.*
3: sp fungi.*
4: sp human.*
5: sp invertebrate.*
6: sp mammal.*
7: sp mhc.*
8: sp organelle.*
9: sp phage.*
10: sp plant.*
11: sp rodent.*
12: sp virus.*
13: sp vertebrate.*
14: sp unclassified.*
15: sp virus.*
16: sp bacteriap.*
17: sp archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

		SUMMARIES			
result	No.	Score	Query Match	Length	Description
1	1156.5	39.9	623	3	Q8J136 penicillina
2	976.5	33.7	564	3	Q8J2V8 tricholoma
3	175.5	6.0	522	2	Q8GAK8 arthroba
4	173.5	6.0	539	2	Q8QEQ7 burkholderi
5	170.5	5.9	573	16	Q9P190 campylobact
6	170	5.9	551	2	Q9RH54 pantoea agg
7	166.5	5.7	579	16	Q9A7T6 caulobacter
8	159	5.5	615	2	O34214 pectobacter
9	158.5	5.5	529	17	Q9HOR8 halobacteri
10	158.5	5.5	619	16	Q8YWM9 Q8YWM9 anabaena sp
11	157	5.4	545	2	Q9LBR8 gluconobact
12	155	5.3	539	16	Q8YLB9 ralstonia s
13	148	5.1	553	2	Q9XCR0 pantoea cit
14	144	5.0	722	16	Q9R226 deinetococcus
15	141	4.9	523	16	Q98C76 rhizobium 1
16	141	4.9	591	16	Q911K8 pseudomonas

17	135	4.7	561	16	Q8U7S2
18	131.5	4.5	528	2	Q8KPV4
19	128.5	4.4	748	10	Q9M0H4
20	128.5	4.4	748	10	Q8LDP0
21	128.5	4.4	748	10	Q94BP3
22	127	4.4	499	16	Q988P2
23	123.5	4.3	508	16	Q8PKP1
24	122	4.2	527	16	Q8U803
25	119	4.1	518	16	Q8EYN5
26	117.5	4.1	620	3	Q12623
27	116.5	4.0	494	16	Q8FY45
28	116.5	4.0	609	2	Q8KIL0
29	116	4.0	502	5	O18672
30	116	4.0	599	5	Q18429
31	114.5	3.9	704	3	Q9P8D9
32	114	3.9	769	3	Q74253
33	113	3.9	1012	10	O64492
34	113.5	3.9	544	16	Q8XU48
35	113.5	3.9	578	16	Q57307
36	112	3.9	768	3	Q8J2T4
37	112	3.9	1881	12	Q9DUN3
38	111.5	3.8	494	16	Q8YJ07
39	111.5	3.8	586	10	Q8H094
40	111	3.8	1492	5	Q8SSU1
41	110.5	3.8	589	10	Q9M4V5
42	110	3.8	537	16	Q8F925
43	110	3.8	551	2	Q8GME7
44	110	3.8	1498	16	Q8YA05
45	109	3.8	418	16	Q8EL08

ALIGNMENTS

RESULT 1

Q8J136 PRELIMINARY; PRT; 623 AA.
AC Q8J136
DT 01-MAR-2003 (TRENBLrel. 23, Created)
DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)
DE 01-MAR-2003 (TRENBLrel. 23, Last annotation update)
DE Pyranose oxidase (EC 1.1.3.10).
OS Penicillina sp. SG.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Aphyllophorales; Lachnolcladiaceae; Penicophora.
OX NCBI_TaxID=204723;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SG;
RA Heckmann D.M., Bastian S., Rekowski M., Giffhorn F.;
RT "Pyranose Oxidase of the White-Rot Fungus Penicophora sp. Strain SG:
RT Cloning and Characterization of the Gene, Heterologous Expression in
RT Escherichia coli, and Properties of the Recombinant Enzyme.";
RL Submitted (AUG-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF535193; AAO13382.1;
KW Oxidoreductase.
SQ SEQUENCE 623 AA; 69342 MW; 4B7DB64A387F43D5 CRC64;

Query Match 39.9%; Score 1156.5; DB 3; Length 623;
Best Local Similarity 45.7%; Pred. No. 7.3e-89;
Matches 253; Conservative 69; Mismatches 159; Indels 73; Gaps 15;
Oy 13 GYHKNEIFQKIDRFVWIKGALOVSVPVRNQNVTLPDPCGWSAPPSAISNGKNP 72
Db 87 GAHKNVVEQKIDRFVWIKGALOVSVPVRNQNVTLPDPCGWSAPPSAISNGKNP 144
Oy 73 HQREFENLSAEATRVGGGMSHTWCTSPRIHPMESLPGLGPKLNDPAE-DDKEWNE 131
Db 145 EQDPLRNLSCQAVTRVVGGMSTHTWCTATPFDREQ-----RPLLKDDADDAEDR 197
Oy 132 LYSEAEERLIGTSTKEFDESRHTLVLSLQDAYKDRQIFRPLPLACHRLKNAPEYWEH 191
Db 198 LYTAKESYFQTGTDQFKESIRHNLNKLAEYKQ-ORDFQOIPLAATR--RSPTFVWS 254

QY 192 SAENLF--HSTYNDKOKKFTLLTNHRTCLALTGGVEKKIGAEVRLATRNPSQL 249
 DB 255 SANTVFDLQNRPNTPDAEERENLFPVACERV-----VRNAL-----NSEI 295
 QY 250 DSY-----IMAKVYVLASGAIGNPQILYNSFGSL-QVTPRN-DSLIPNLGRY 295
 DB 296 ESLHIDLISGRPEIKADVVVLTAGAVHNTQLLVNSFGQLGRPNPTNPPELLPSLSY 355
 QY 296 ITEOPMAFCQIVLRQEFVDSVRDPP--YGLP-----WKEAQAQH 333
 DB 356 ITEOSLFCQVMSTELISYKSDMTIRGTPGELTYSVITPGASTNKHKPDWNEKVKH 415
 QY 334 IAKNPTDALPTPRDPDPQVTPTEBHPWHTQIHRDAFSGVAGVPRVDSRVLDLRWFG 393
 DB 416 MMQOEDEPLIPFEDPEPQVITLFPQSPHWTQIHRDAFSGVAGVQOISDRLLVDNRFFG 475
 QY 394 ATDPEANLLVFNQDVODGYSMPQPTFRYR-PTASNVRRARKMADMCVSNLGGYLPT 452
 DB 476 RTEPKENKLFWSKIIDAYNMPOPTDFRFPAGRTSKEAEDMTDMCVMSAKIGGFLPG 535
 QY 453 SPQFMDPGLALHLAGTTRGFDA--TTVADNNSLVWDFANLYVAGNGTIRTGGENPT 510
 DB 536 SLPOFMEPGLVLLGGTHRMGFDEKEDNCCVNTDSRVFGFKNLFGLGCGNIPYAGNPT 595
 QY 511 LTSMCHAIKSARSI 524
 DB 596 LTMSLAIKSCEYI 609
 RESULT 2
 Q8J2V8
 ID Q8J2V8 PRELIMINARY; PRT; 564 AA.
 AC Q8J2V8;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Pyranose 2-oxidase (EC 1.1.3.10).
 GN P20.
 OS Tricholoma matsutake.
 OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
 OC Agaricales; Tricholomataceae; Tricholoma.
 CX NCBI_TaxID=40145;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Takakura Y.;
 RT "Purification, characterization and primary structure of a pyranose
 RT oxidase from Tricholoma matsutake."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB043883; BAC24805.1;
 KW Oxidoreductase;
 SQ SEQUENCE 564 AA; 61942 MW; 134790030045FCIB CRC64;
 Query Match 33.7%; Score 976.5; DB 3; Length 564;
 Best Local Similarity 41.3%; Pred. No. 1e-73;
 Matches 223; Conservative 82; Mismatches 182; Indels 53; Gaps 13;
 QY 2 AEGTAVPYPGTHKKEIEFQKIDRPNVNIKALQOQVSVVRNQNVTLPDPAWSAPP 61
 DB 62 AEIGSQDNPIVGAHRRNSIKFQKIDKFVNIINGALQIPISPSDYQPTLVAVAAM-APP 120
 QY 62 GSSA----ISNGKNPHOREFENISAEAVTRGVGSMTHWTCSPTRTHPPMESLPGIGRPK 117
 DB 121 IDPAEGQLVINGHNPQEAENLPGSATVTVGGMATHWTCACPTPHD-----EE 170
 QY 118 LSNDAEDDKEMNLYSEAEIRLTGTKEFDESIRHTLVLSQDAYKQRIFRPLPLA 177
 DB 171 RVNPNV-DKQEPDALLERAKTLNVHSDQYDDSIQIVKVKETLQOQL-DASRGVTLPLG 228
 QY 178 CHLKNAPVEVHSHENLPHSYINDDKQKQLFTLLTNHRTCLALTGGVEKKIGAEVR 237
 DB 229 VERTDNPIVTVTGAD----TVLGDVPKSPREVLVTETRVTKFVIVSETNPTQVVAALLR 284

QY 238 NLLATRNPSQLDSYIMAKVYVLASGAIGNPQILYNSFGSLQVTPRNDSLIPNLGRYIT 297
 DB 285 NL-----NTSN--DELVAQSFVIACGAVCTPQLWNS-----NIRH-----ALGRYLS 327
 QY 298 EQPMAPFCQIVLRQEFVDSVRDPPYGLPWWKEAQAQIAKNPTDALPFPDRPDPQVTPPF 357
 DB 328 EQSMTFCQIVLKRISIVDSIATDPR-----FAAKVEAHKKKHPDDVLPIPPHEPSPQVMIPY 383
 QY 358 TEEHPWHTQIHRDAFSGVAGVPRVDSRVLDLRWFGCATDPEANLLVFP-----QND 408
 DB 384 TSDFPWHVQVHR--YAFGDVGPKADPRVVVDLRFKGSQDIVENRVTFGNPNKLRDWEAG 441
 QY 409 VODGYSMPQPTFRYR-PTASNVRRARKMADMCVSNLGGYLPTSPQFMDPGLALHLAG 469
 DB 442 VDTYGMPOPTFHVKTNDAGDQDQRMNDMTNVANILGGYLPSPQFMAPGLAQHTG 501
 QY 469 TTRIGFDKATTVADNNSLVWDFANLYVAGNGTIRTGGENPTLTSCHAIKSARSIINTL 528
 DB 502 TTRIGTDDQTSVADPTSKVHNFNVLWVGNGCIPDATACTNPTRTSVAYALKGAEEVSYL 561
 RESULT 3
 Q8GAK8
 ID Q8GAK8 PRELIMINARY; PRT; 522 AA.
 AC Q8GAK8;
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE 2-keto-glucuronate dehydrogenase-like protein.
 GN KDG.
 OS Arthrobacter nicotinovorans.
 OC Plasmid PA01.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Micrococccineae; Micrococcaceae; Arthrobacter.
 CX NCBI_TaxID=29320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95115562; PubMed=7815950;
 RA Grether-Beck S., Igloi G.L., Pust S., Schiltz E., Decker K.,
 RA Brandsch R.;
 RT "Structural analysis and molybdenum-dependent expression of the PA01-
 RT encoded nicotine dehydrogenase genes of Arthrobacter nicotinovorans";
 RL Mol. Microbiol. 13:929-936(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96172783; PubMed=8588735;
 RA Menendez C., Igloi G., Henninger H., Brandsch R.;
 RA "A PA01-encoded molybdopterine cofactor gene (moaA) of Arthrobacter
 RT nicotinovorans: characterization and site-directed mutagenesis of the
 RT encoded protein";
 RL Arch. Microbiol. 164:142-151(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97230479; PubMed=9073590;
 RA Menendez C., Igloi G.L., Brandsch R.;
 RT "IS1473, a putative insertion sequence identified in the plasmid PA01
 RT from Arthrobacter nicotinovorans: isolation, characterisation and
 RT distribution among Arthrobacter species";
 RL Plasmid 37:35-41(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98088982; PubMed=9428706;
 RA Menendez C., Otto A., Igloi G., Nick P., Brandsch R., Schubach B.,
 RA Botcher B., Brandsch R.;
 RT "Molybdate-uptake genes and molybdopterine-biosynthesis genes on a
 RT bacterial plasmid. Characterization of MoaA as a filament-forming
 RT protein with adenosinetriphosphatase activity";
 RL Eur. J. Biochem. 250:524-531(1997).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Schenk S., Hoelz A., Kraus B., Decker K.;
 RT "Gene structure and properties of enzymes of the plasmid-encoded
 RT nicotine catabolism of Arthrobacter nicotinovorans";

J. Mol. Biol. 284:1323-1339(1999).
[6]
SEQUENCE FROM N.A.
MEDLINE=21405725; PubMed=11514508;
Baitsch D., Sandu C., Brandsch R., Igloi G.L.;
"A gene cluster on pAOI of *Arthrobacter nicotinovorans* involved in the
degradation of the plant alkaloid nicotine: cloning, purification and
characterization of 2,6-dihydroxypyridine 3-hydroxylase".
J. Bacteriol. 183:5262-5267(2001).
[7]
SEQUENCE FROM N.A.
Igloi G.L., Brandsch R.;
"Sequence of the 165 kb Catabolic Plasmid pAOI from *Arthrobacter*
nicotinovorans and Identification of a pAOI-dependent Nicotine Uptake
System".
Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
JL EMBL: AJ507836; CAD47896.1; -;
JW Plasmid.
JQ SEQUENCE 522 AA; 56526 MW; ADB9B35DB3CCC939 CRC64;

Query Match 6.0%; Score 175.5; DB 2; Length 522;
Best Local Similarity 21.8%; Pred. No. 5.2e-06;
Matches 121; Conservative 58; Mismatches 208; Indels 167; Gaps 24;

Y 27 DRFVNVKALQOVSVVRNQNVTLDPCWAGAP---PGSSALSNGKNPHQREFENLSAE 83
b 47 DXFINDERESYKLT-----WTDTRLATGWSLAKDFFGSPA-WNGK----- 87
Y 84 AVTRGVGGMSTHTCSTPRI-----HPPMESLPGIGRKLSPNDPAEDDKENNELYSEA 136
b 88 ---AVGGTATETWGLTPFKWHEFKTHYYGLPD---GTIADWPLDLD-ELDHYHTAA 139
Y 137 ERLICTSKPEDESHTVLVLSLQDAYKQRIIFRPLPLACHRLKNAPEYVEM-HSAEN 195
b 140 EKAVGASH-----RHGRPPLPASNGYK-----VLANGAERIGYRHYATG 178
Y 196 LFHS-----IYNDKQKXKLLTLLTNHRCRLALTGGEKKIGAEVRNL 239
b 179 PYATNVEYDGRPGTVQDGFAMAGDSRAKWSPLVSEIPKALA-TGLLELRTEAQAQV-- 235
Y 240 LATRNPPSOLDSYI-----MAKYVVLASGAIGNPQILYNSGSLQVTPRNDSL 288
b 236 -ITLPGDGRAGVVYTDVGNLQGRARLVAVAGNAIETPRLLLSATSG-----HTDGL 289
Y 289 -----IPNLGYITEQPMACQIVLROBFVDSVRDDP---YGL 323
b 290 ANSSGQVGRNMYRHTTGLVYAEFFNEVHMVYRGEPMAGI-----ISDSRHPDPSRGFVG 342
Y 324 PMWKEAVAQHIKNTDAPLPIPRDPEQVTPTEHPHWTQIHRDAFSVGAVGPEVDS 383
b 343 GYIEMIAQGLPSFT-----FMSP-----GEWGPQFTE 371
Y 384 RVIVDLR-----WFGATD-PEANLLVFONDQDGYSMPOPTFRYPSTASNRARKQWAD 438
b 372 KVEAYTRTAALWICGEDVDPQASNRVTLNLTVDRLGLPAPVHY-DHPNDIAHRNHGYE 430
Y 439 MCEVASNLGGVLPSPQFMDPGALHLAGTTRIGFKATTVADNNSLVWDFANLYVAGN 498
b 431 QGEKLFKSVGAIRTRAPGMSG---HNLGTARMSQKPDGCVNSFGQAHVDPNLFVSDG 487
Y 499 GTIRTFGENPTLT 512
b 488 SOFTTGAANPTLT 501

RESULT 4
Q9Q67 PRELIMINARY; PRT; 539 AA.
Q9Q67;
01-MAR-2003 (TrEMBLrel. 23, Created)
01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
Glucose dehydrogenase.

GN GDHALPHA.
OS Burkholderia cepacia (pseudomonas cepacia).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OX NCBI_TaxID=292;
RN [1]
RP SEQUENCE FROM N.A.
RA Sode K., Inose K., Fujikawa M., Yamazaki T.;
RT "Molecular cloning of the gene encoding the most stable glucose
dehydrogenase from a moderate thermophilic bacterium".
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF430844; AAN39686.1; -;
SQ SEQUENCE 539 AA; 59832 MW; 9CB4A4E5D39AE845 CRC64;

Query Match 6.0%; Score 173.5; DB 2; Length 539;
Best Local Similarity 19.9%; Pred. No. 8e-06;
Matches 109; Conservative 80; Mismatches 191; Indels 169; Gaps 24;

QY 60 PGSSAISNGKNPHQREFENLSAEAVTRGVGGMSTHTCSTPRIHP---PMESLPGIGRP 116
b 78 PPNLYLILKGEHKFNSQV-----IRAVGGTTWHWAASAMRFIPNDFKMSVYGVGR- 128
QY 117 KLSNDPAE-DDKENNELYSEAEERLIGTSTKEPDESIRHTLVLSLQDAYKQRIIP--RP 173
b 129 ---DWPIQYDLE--PYQRAEEELGVMGPGPE-----DLYSPKQPYMPMP 171
QY 174 LPLACHR--LKNAPYEVWHSANL-FHSI-----YNDKQKXKLLTLLTNHRCRLALTOG 226
b 172 LPLSFNEQITKA-----LNNYDPKFXVVTBPARNSRPYDGRPTCCGNNNCMPICPGA 226
QY 227 Y-----EKKTGAAEVRNLATR---NPSQL-----DSYIMAKVYVLASG 263
b 227 MVNGIVHVEKAERAGAKLIENAVVYKLETPDKRIVAALYKDKTGAHRVEGKYVFLAAN 286
QY 264 AIGNPQILYNSGSLQVTPRNDSLIPNLGYITEQPMACQIVLROBFVDSVRDDPYGL 323
b 287 GIETPKIL-----LMSANRDFPNGV 306
QY 324 PMWKEAVAQHIKNTDAPLPI-----PFRDPEQVTPTEHPHWT-----QIH--- 368
b 307 ANSDMVGRNLMHPGTGVSFASEKLMWPGRPOBMTSLIGFRDGPFRATEAARKIHLSN 366
QY 369 -----RDAPSYGAV--GPEVDSRV-----TVDLRFQATDPEANLLVFPQNDVODG 412
b 367 LSRIDQETQKIFKAGKMKPDELDAQIRDSARYVQDFCFHEILPOENRIVPSKTATDA 426
QY 413 YSMPOPTFRYPSTASNRARKQWADMEV---ASNLGG-----YLPSPQFMDP 460
b 427 IGIPRPEITY-----AIDDYVKRGAHTREVYATAAKVLGGTDVVVDFEAFNN----- 475
QY 461 GLALHLAGTTRIGFKATTVADNNSLVWDFANLYVAGNRTGTGFPENPTLTSCHAIKS 520
b 476 ---HITGTINGADARDSVVDKCRTFDHPNLFSSATMPTVGTNVTLTIALALRM 531
QY 521 ARSIINTLK 529
b 532 S---DTLK 536

RESULT 5
Q9PI90 PRELIMINARY; PRT; 573 AA.
Q9PI90;
AC Q9PI90;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE Putative oxidoreductase subunit.
GN CJO415.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
OC Campylobacteraceae; Campylobacter.
OX NCBI_TaxID=197;
RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kelsey J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Meltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrall B.G.;
 RT "The genome sequence of the food-borne pathogen *Campylobacter jejuni*
 RT reveals hypervariable sequences.";
 RL Nature 403:668-668(2000).
 DR EMBL; AL139075; CAB74251.1; -;
 KW Complete proteome.
 SQ SEQUENCE 573 AA; 63686 MW; 7BEB522A779A5F94 CRC64;

Query Match 5.9%; Score 170.5; DB 16; Length 573;
 Best Local Similarity 20.9%; Pred. No. 1.6e-05;
 Matches 111; Conservative 61; Mismatches 237; Indels 123; Gaps 21;

QY 77 FENLSAEAVTRGVGGMSTHTCSTPRTHPPWESLPGIGRP-----KLSNDPAEDD-----K 127
 DB 84 YRMGSLGNNVGGAGVHNGWTFRPMYPDFPIQLSKORYGNKLGNDYTLQDWGVYK 143
 QY 128 EWNELYSERLIGTSTKEF-----DESIHRTLVLRSLODAYKQRIER 172
 DB 144 DMEPYDRFEKTCGVSGEPNLAEMKGAFRSSPYQPEPLENTKMLKRFESAASSNLHTY 203
 QY 173 PLPLA-----CHRLKNAPEYVEHSAENLFHSIYNDKQKLF 210
 DB 204 RLPASNSKGGYTNPDGODLAPQVCAICERFG-----CEYGAKASPLNTVIPKAMSTGKY 258
 QY 211 TLLTNHRTCTALTGVEKKIGAEAVRNLLATRNPSQLDSYIM-AKVYVLSAIGNQP 269
 DB 259 TIRYSNVTLQ-----KKQKTVGVKFDVTRT-----MKEYIQPADIVVLTYSMNNAK 308
 QY 270 ILYNSGFGSLQVTP-----RNDSLIPNLG--RYITEQMAFCQVLRQEFVDSVRDD 319
 DB 309 LLAVSNI-GRQYDKTKGTLGRNYCYQMMGTATAFDEQFNTP-----MSSGALGTTSD 363
 QY 320 PYG--LPWKE-----AVAQIAKNTDALLIPFRDPPEQVTPPTTEHPHWTQIHRD 370
 DB 364 FNGDNFDSKEFLHGAMIYSVOLGTRPIQSAFLPAGAP-----TWGAEF-KK 410
 QY 371 AFSYG-----AVGPEVDSRVIVDLRFEGATDPEANLLVFQNDVQDGYSMQPTFRYRS 425
 DB 411 ALNTNFRATVGGQ-----GASLPHKNYLSLDPTYKDAFGMLRLTYN-F 457
 QY 426 TASNVRARKMAD-MCEVASNLGGLYPTSPQFMD-----PGLALHLAGTTRIGFDKATT 479
 DB 458 TDQDRALHKEFITDKTAEAVKRMQGVKSIKGAVALKDYVVVYQSTHTGTTNGADRETS 517
 QY 480 VADNNSLVWDFANLYVAGNGTIRTGFGENTLTSMCHAIKSARSINTLKG 531
 DB 518 VVNTYLQHWADNLFFVVGAGNFQHNSGYNPTDTVGALAYRCAEGILKYHSG 569

RESULT 6

Q9RH54 PRELIMINARY; PRT; 551 AA.
 AC Q9RH54;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE Dehydrogenase subunit 1.
 OS Pantoea agglomerans.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 OC Enterobacteriaceae; Pantoea.
 OX NCBI_TaxID=549;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chang K.-I., Yum D.-Y., Pan J.-G., Shin Y.-C.;
 RT "Cloning and expression of a gene cluster encoding three subunits of
 RT membrane-bound 2-keto-D-gluconate dehydrogenase from *Erwinia herbicola*

RT ATCC08111 in *Escherichia coli*.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF068066; AAF21261.1; -;
 DR InterPro; IPR00205; NAD binding.
 SQ SEQUENCE 551 AA; 60047 MW; FE2ECCD0933DAFBF CRC64;

Query Match 5.9%; Score 170; DB 2; Length 551;
 Best Local Similarity 20.3%; Pred. No. 1.6e-05;
 Matches 109; Conservative 65; Mismatches 209; Indels 154; Gaps 21;

QY 68 NGKNPHOREFENLSAEAVTRGVGGMSTHTCSTPRTHPP-----MESLPGIGRPLKSLNDPAE 124
 DB 93 NVTGPNADSFQ-----QGYLRTVGGTTHWAASCRHHPSPDFVMSQYGVGR----- 139
 QY 125 DDKEM-----NEL---YSEAEERLIGTSTKEFDESIHRTLVLRSLODAYKQRIERPLA 177
 DB 140 ---DWPIGVDELEPWYCKAENEIGVAGP--NDPARQSPTERS-----QPYPM- 181
 QY 178 CHRLKNAPEYVEHSAENLFHSIYNDKQKLF-----FTLLTNHRTCTAL 223
 DB 182 -----DMVPPAHGDNYPASVVPNGYNLVPFIPQGRSTRPWEGRPTCCGNNNCQFICP 233
 QY 224 TGYEKKIGAAEVRNLLATRNPSQLDSYIM-----AK 256
 DB 234 IGAMTYGIHVE-----RAERNGAVVLAABAVVYKMDTDSNNRITAVHMLDTSGASHKATAK 289
 QY 257 VVVLASGALGNQILYNSGFGSLQVTPRNDSLIPN---LGRVITEQMAFCQVLRQEF 312
 DB 290 AFALACNGIETRLLL-----MAANDANPNGIANASDMVGRNMDSHGFCSELTKE-- 341
 QY 313 VDSVRDDPYGLPWKEAVAQIAKNT-DALPIFRDPPEQVTPPTTEHPHWTQI----- 367
 DB 342 -----PW-----LCKGPAQSSCMVGYRDG-----FRDYSANKVILNNI 377
 QY 368 -----HRDAFSGVAGVPEVDSRV-----IVDLRFEGATDPEANLLVFQNDVQDGYSM 415
 DB 378 SRVVTATQAMKGLVGLKALDEIRYRAVHSVDLSISLEPLDPENRLTLSTRKDPHGL 437
 QY 416 PPTFRYRSTASNVARKMADCEVASNLGGLYPTSPQFMDPGLAL--HLAGTTRIG 473
 DB 438 PCPDIYDVGVYRKGAESAHAQLEHI-----GQLFDAKEFTISQGLNANNHMGVING 492
 QY 474 FQKATTVADNNSLVWDFANLYVAGNGTIRTGFGENTLTSMCHAIKSARSINTLKG 530
 DB 493 KNAKEAVVQNCCEAFPHENLWLPGGGAI PSASVVNSTLTMAALGLKAAHDI SLRMKG 549

RESULT 7
 Q9A7T6 PRELIMINARY; PRT; 579 AA.
 AC Q9A7T6;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Oxidoreductase, GMC family.
 GN CC1634.
 OS Caulobacter crescentus.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
 OC Caulobacteraceae; Caulobacter.
 OX NCBI_TaxID=155892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 19089 / CB15;
 RX MEDLINE=21173698; PubMed=11259647;
 RA Nierman W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
 RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
 RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
 RA DeBoy R.T., Dodson R.J., Durkin A.S., Ghinn M.L., Haft D.H.,
 RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
 RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
 RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
 RT "Complete genome sequence of *Caulobacter crescentus*.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).

[illegible]

Query Match	5.5%	Score 159;	DB 2;	Length 615;
Best Local Similarity	19.9%;	Pred. No. 0.00017;		
Matches 124;	Conservative 85;	Mismatches 257;	Indels 158;	Gaps 32
QY	4	EGTAVPVYVGGYHKKNEIEFFKQIDIRFVNV1	-----KGALQOVSVVPRNQNVPTLDPGAMS	58
Db	5	ERVSVV-VSGYSRGEGVTANELKKVDVAVVVGFWAGAIMAKELTEAGLNVALYERG	-----60	
QY	59	APPGSSAISNGKNP	-----HQREFENLSAEAVT	87
Db	61	--PHRDTYPCGAYPQSIDELTYNIRKKLFQDLSKSTVTIRHDSQATVAPYPQLAAFLPGT	118	
QY	88	GVGMSHTWTCSTPRIHPMESLPQ	-----IGRPKLSNDPAEDD--KEWNEI--YSAE	137
Db	119	CTGAGLHWSGVHFRVDPVELNLSHVEARYGKNFIPEGMTIQDFGVSYNLEPFDDQAE	178	
QY	138	RLIGTSTKEFDESIRHTLVLSRLQDAY	-----KDRQRIERPLP-----LACHRLKNAPEYVEW	190
Db	179	KVFGTSGSAW--TIKGKMKIGKEGNGFYAPDRSSDF	-----PLPAQKRTYSAQLPAQAAESVGY	235
QY	191	H-----SAENLPHSIYNDD	-----KOKKCLFTL	212
Db	236	HPYDMP SANTSGPYNTYGAQMGFCGCGYACMYSKASPNVNLPALRQEPKFEEL	295	
QY	213	LTNHRCTRLLATGGYEKKIGA	-----AEVRMLLATRNPSSO--LDSYIMAKYVVLASGAIG	266
Db	296	RNNAYVLRLVNLTDGDKRATGVYLDGGQREV	-----QPADLVILSAFQPHNVHMLLSGIG	352
QY	267	NP-QILVNSGFGSLQVTPRNDLSLIPNL	-----GRYITBQPM--AFCOIVLRQEFVDSVRDDPY	321
Db	353	QPNPIINEGVGNRFAYQXNISTUKALFDKNTTNPIFGAGGAGVAVDDFNADNFDHGYP	412	
QY	322	GL-----PWWKEAVAQHTAKNPDTALPIPRDPPEPQVTPPTTEHPMHTQIHRDAFSYGAV	377	
Db	413	GFVGSGSPFW--VNOAGTKPVSGLPTPKGTEN	-----WGSQMK-----AAV	450
QY	378	GPEVDSRVIVDLRWFGATDEANNLLVFQNDVQDGYSP	-----QPTFRVPRSTASNVUR--ARK	434
Db	451	ADTYNNHHISMDAH--GAHQSYRANLYLDLDENYKNVYQPLLRMTFDWQD	-----NDIRMAQF	505
QY	435	MMADMCEV--ASN-----LCGYLPTSPQFMDPGU	-----ALHLAGTTRIGFDKATTVADNNSL	486
Db	506	MVGKRRKITAMPKMIIGG--AKGPOTHEFTTVYQTHMSGGAIMGEDPKTSAVNRYLQ	563	
QY	487	WDFPANLYVAGNGTIRTGFGENPT	510	
Db	564	SWDVPNVFVPGASAFPOGLGYNPT	587	
RESULT 9				
Q9HQH8		PRELIMINARY;	PRT;	529 AA.
AC	Q9HQH8			
ID	01-MAR-2001 (T-EMBLrel. 16, Created)			
DT	01-MAR-2001 (T-EMBLrel. 16, Last sequence update)			
DT	01-MAR-2003 (T-EMBLrel. 23, Last annotation update)			
DE	Vng1035C.			
GN	VNG1035C.			
OS	Halobacterium sp. (strain NRC-1)			
OC	Archaea; Euryarchaeota; Halobacteriia; Halobacteriales;			
OC	Halobacteriaceae; Halobacterium.			
OX	NCBI_TaxID=64091;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20504483; PubMed=11016950;			
RA	NG W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,			
RA	Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,			
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,			
RA	Leithausen B., Keller K., Cruz R., Danson M.J., Hough D.W.,			
RA	Maddocks D.G., Jablonksi P.E., Krebs M.P., Angevine C.M., Dale H.,			
RA	IseNBarger T.A., Peck R.F., Pohlshroder M., Spudich J.L., Jung K.-H.,			
RA	Ajam M., Freitas T., Hou S., Daniels C.J., Dennis P., Omer A.D.,			

RP SEQUENCE FROM N.A.
RC STRAIN=IFO 3254;
RA Yoshikawa K., Saito Y., Iehii Y., Neguchi Y., Soeda S.;
RT "PCT Patent Publication No. WO99/0763A1";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB039821; BAA92378.1; -;
DR InterPro: IPR000172; GAC oxid.
DR InterPro: IPR000205; NAD binding.
DR Pfam: PF00732; GMC oxid. 1.
SQ SEQUENCE 545 AA; 60076 MW; 4C0A4F3C56950E22 CRC64;

Query Match 5.4%; Score 157; DB 2; Length 545;
Best Local Similarity 19.7%; Pred. No. 0.0002;
Matches 116; Conservative 69; Mismatches 198; Indels 206; Gaps 25;

QY 34 KGALQVSVPRNQNVPTLDPGAWSPGSAISNG-----KNPHREFENLSAEAVTRGV 89
DB 34 KGALQVLPYPPV-----PWAMHPDQGSFNGYLHTTGPDAAYQ---QGVLVV 102

QY 90 GGMSTHTCTSTPRIHP---PMESLPGIGRKLNDPAEDDKENNELYSEARLIGTSTKE 146
DB 90 GGMSTHTCTSTPRIHP---PMESLPGIGRKLNDPAEDDKENNELYSEARLIGTSTKE 146

QY 103 GGTTHWACARWYLPDSFELSHRYGVR-----DWAIKYDDLEPPYQAEVMMVAGEN 157
DB 103 GGTTHWACARWYLPDSFELSHRYGVR-----DWAIKYDDLEPPYQAEVMMVAGEN 157

QY 147 FDESIRHTLVLSLODAYKDRQIF--RLPLACHRLKNAPEYVWHSANLPHSIYND 204
DB 147 FDESIRHTLVLSLODAYKDRQIF--RLPLACHRLKNAPEYVWHSANLPHSIYND 204

QY 158 MD-----VDDLGSFGRSHNYPKKEVPLS-----YGAD 183
DB 158 MD-----VDDLGSFGRSHNYPKKEVPLS-----YGAD 183

QY 205 KKKKLTLLTNHR-----CTRLALTGG-----YEKKI 231
DB 205 KKKKLTLLTNHR-----CTRLALTGG-----YEKKI 231

QY 184 OPRKLIHEKTYRNVVHVEPOARNTRPYDKRPTCEGNNNCMPICPIGAMYNGIHSVNAEAA 243
DB 184 OPRKLIHEKTYRNVVHVEPOARNTRPYDKRPTCEGNNNCMPICPIGAMYNGIHSVNAEAA 243

QY 232 GAAEVRNLLATNPSSQSDSYM-----AKVYVLAGAT----- 265
DB 232 GAAEVRNLLATNPSSQSDSYM-----AKVYVLAGAT----- 265

QY 244 GARIIPNAVYRLETDASKNKYVPVNYDPPDKNSHRVTGKFFVAAHCIESAKLLLSAD 303
DB 244 GARIIPNAVYRLETDASKNKYVPVNYDPPDKNSHRVTGKFFVAAHCIESAKLLLSAD 303

QY 266 -GNPQILYNSG-----PSGLQV--PRNDSLIPNLGRITVQPMACFQIVLRQEPV 313
DB 266 -GNPQILYNSG-----PSGLQV--PRNDSLIPNLGRITVQPMACFQIVLRQEPV 313

QY 304 DKNPFIANSDDQVGNMMDHGTGVLSPNSGNSLWPGRGPLT-----SII 350
DB 304 DKNPFIANSDDQVGNMMDHGTGVLSPNSGNSLWPGRGPLT-----SII 350

QY 314 DSVRDDPVLPMWKE--VAQHIKNTDALTPIPRDPPEQVTPPTTEHPWHTQIHRDA 371
DB 314 DSVRDDPVLPMWKE--VAQHIKNTDALTPIPRDPPEQVTPPTTEHPWHTQIHRDA 371

QY 351 DSFRDG----PWSRSGAYLVHMVD-----DNQVDFATGL-----A 382
DB 351 DSFRDG----PWSRSGAYLVHMVD-----DNQVDFATGL-----A 382

QY 372 FSYGAVGPEVDSRVLD-----LRWF-----GATDPANNLLVQNDVQGSNMPQTFYR 423
DB 372 FSYGAVGPEVDSRVLD-----LRWF-----GATDPANNLLVQNDVQGSNMPQTFYR 423

QY 383 IAKGVGKELEEQIRYGSASHVRLFSHNEGIADPD--NRLTLSTKHKDVLGIPHPVEYVK 440
DB 383 IAKGVGKELEEQIRYGSASHVRLFSHNEGIADPD--NRLTLSTKHKDVLGIPHPVEYVK 440

QY 424 ---PSTASNVRAKKNMADCEVASNLG-----GYLPTSPQFMDPGLALHLAGITRIGFD 475
DB 424 ---PSTASNVRAKKNMADCEVASNLG-----GYLPTSPQFMDPGLALHLAGITRIGFD 475

QY 441 LPEYTVKSCDHTKELFKELMALMSGTDPOWTKGYPP-----OCHPSGSGTIMGTD 489
DB 441 LPEYTVKSCDHTKELFKELMALMSGTDPOWTKGYPP-----OCHPSGSGTIMGTD 489

QY 476 KATTVADNNSLVDFANLYVAGNGTIRGTGFGENPTLTSCHAIKASRSI 524
DB 476 KATTVADNNSLVDFANLYVAGNGTIRGTGFGENPTLTSCHAIKASRSI 524

QY 490 PTNSVVDGECRTHDHNLFVARSAVFSVGTGNTLTIGALALRAVLAASL 538
DB 490 PTNSVVDGECRTHDHNLFVARSAVFSVGTGNTLTIGALALRAVLAASL 538

RESULT 12
ID Q8Y1B9 PRELIMINARY; PRT; 539 AA.
C Q8Y1B9;
T 01-MAR-2002 (TrEMBLrel. 20, Created)
T 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
T 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
E Putative transmembrane dehydrogenase (Large subunit) oxidoreductase
F protein (EC 1.1.1.1).
F RSC0771 OR RS05083
F Ralstonia solanacearum (Pseudomonas solanacearum).
F Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
F Ralstoniaceae; Ralstonia.
F NCBI_TaxID=305;
F [1]
F SEQUENCE FROM N.A.
F STRAIN=GMI1000;

RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choine N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Nature 415:497-502(2002).
DR EMBL: AL646061; CAD14473.1; -;
DR InterPro: IPR000566; Lipocin cyfBAP.
DR InterPro: IPR000205; NAD binding.
DR InterPro: IPR001100; Pyr redox.
DR PRINTS: PR00411; PNRDRTASEI.
DR PROSITE: PS00213; LIPOCALIN; 1.
DR Oxidoreductase; Complete proteome.
SQ SEQUENCE 539 AA; 59680 MW; 2F848DA6F0B6F0E7 CRC64;

Query Match 5.3%; Score 155; DB 16; Length 539;
Best Local Similarity 20.7%; Pred. No. 0.00029;
Matches 107; Conservative 72; Mismatches 219; Indels 120; Gaps 21;

QY 60 PGSSAISNGKPHOREPENLSAEAVTRGVGMSTHTCTSTPRIHP---PMESLPGIGRP 116
DB 60 PGSSAISNGKPHOREPENLSAEAVTRGVGMSTHTCTSTPRIHP---PMESLPGIGRP 116

QY 79 PDNNYLLIKGEHKFDSQY-----IRAVGGTTWHAASTWRFLPNDPKLSRVYGIAR- 129
DB 79 PDNNYLLIKGEHKFDSQY-----IRAVGGTTWHAASTWRFLPNDPKLSRVYGIAR- 129

QY 117 KLSNDPAEDDKENNELYSEARLIGTSTKEP-----DESI----- 151
DB 117 KLSNDPAEDDKENNELYSEARLIGTSTKEP-----DESI----- 151

QY 130 -----DWPQYDDLERYGREAALGVMGFNDSEELGSPRSQPYMAPLPLSPNE 178
DB 130 -----DWPQYDDLERYGREAALGVMGFNDSEELGSPRSQPYMAPLPLSPNE 178

QY 152 -----RHTLVLSLODAYKDRQIFRPLPLACHRLKNAPEYVWHSANLPHSIYND 203
DB 152 -----RHTLVLSLODAYKDRQIFRPLPLACHRLKNAPEYVWHSANLPHSIYND 203

QY 179 RTVKDALNAHPALRVVTEPVARSRPYDGRPTCCGNNNCMPIC-----PIGAMYNGI 234
DB 179 RTVKDALNAHPALRVVTEPVARSRPYDGRPTCCGNNNCMPIC-----PIGAMYNGI 234

QY 204 DKQKLLFTLLNHRCTRALTGGYEKKIAGAEVRNLLATNPSSQSDSYMAMKVYVLAGS 263
DB 204 DKQKLLFTLLNHRCTRALTGGYEKKIAGAEVRNLLATNPSSQSDSYMAMKVYVLAGS 263

QY 235 EKAQAGARLIENAVV-----YKLEVGAGK-RIVAAHYKDPKGVDRHVEGKWFVLAAN 286
DB 235 EKAQAGARLIENAVV-----YKLEVGAGK-RIVAAHYKDPKGVDRHVEGKWFVLAAN 286

QY 264 AIGNPOILYNSGFSGLQ--VTPRNDSLIPNLGRITVQPMACFQIVLRQEFVDSVRDDPY 321
DB 264 AIGNPOILYNSGFSGLQ--VTPRNDSLIPNLGRITVQPMACFQIVLRQEFVDSVRDDPY 321

QY 287 GIETPKMLMSTGADFPKGVGNRSDM-----VGRNLMHPGTGVSFYADKL----- 333
DB 287 GIETPKMLMSTGADFPKGVGNRSDM-----VGRNLMHPGTGVSFYADKL----- 333

QY 322 GLPWKEAQAHIKNTDALTPIPRDPPEQVTPPTTEHPWH-TQIHRDA---PSYGAV 377
DB 322 GLPWKEAQAHIKNTDALTPIPRDPPEQVTPPTTEHPWH-TQIHRDA---PSYGAV 377

QY 334 ---WPGEGPEMTS-----LIGFRGPFRTAQGKLLHLSNISRIEQETTRIFKAGKL 383
DB 334 ---WPGEGPEMTS-----LIGFRGPFRTAQGKLLHLSNISRIEQETTRIFKAGKL 383

QY 378 --GPEVDSRV-----IVDLRFWFGATDPEANNLLVQNDVQGSNMPQTFYRPS---T 426
DB 378 --GPEVDSRV-----IVDLRFWFGATDPEANNLLVQNDVQGSNMPQTFYRPS---T 426

QY 384 LKPAELDARIRDOAARYVQDPSFHEILLPLPENRVPSATGTDALGIRPEITYRIDYVK 443
DB 384 LKPAELDARIRDOAARYVQDPSFHEILLPLPENRVPSATGTDALGIRPEITYRIDYVK 443

QY 427 ASNVRAKKNMADCEVASNLGGLYLPSPQFMD---PGLALHLAGITRIGFDKATTVADN 483
DB 427 ASNVRAKKNMADCEVASNLGGLYLPSPQFMD---PGLALHLAGITRIGFDKATTVADN 483

QY 444 RSAVHTREVYATAARV---LGG-----TDVQFHDHDFAPN--NHITGATIMGSDPKDSVVDK 494
DB 444 RSAVHTREVYATAARV---LGG-----TDVQFHDHDFAPN--NHITGATIMGSDPKDSVVDK 494

QY 484 NSLVDFANLYVAGNGTIRGTGFGENPTLTSCHAIKAS 521
DB 484 NSLVDFANLYVAGNGTIRGTGFGENPTLTSCHAIKAS 521

QY 495 DCRTFDHPNLFISSATMPVGTGVNVTLTIAALRIA 532
DB 495 DCRTFDHPNLFISSATMPVGTGVNVTLTIAALRIA 532

RESULT 13
ID Q9XCRO PRELIMINARY; PRT; 553 AA.
C Q9XCRO;
T 01-NOV-1999 (TrEMBLrel. 12, Created)
T 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
T 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
E 2-keto-gluconate dehydrogenase subunit.
F K0GB.
F Pantoea citrea.
F Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
F Enterobacteriaceae; Pantoea.
F NCBI_TaxID=53336;
F [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=1056R;
 RX MEDLINE=20200361; PubMed=10735866;
 RA Pujol C.J., Kado C.I.;
 RT "Genetic and Biochemical characterization of the pathway in Pantoea
 citrea leading to pink disease of pineapple.";
 RL J. Bacteriol. 182:2230-2237(2000).
 DR EMBL; AF131202; AAD44706.1; -;
 DR InterPro; IPR000205; NAD_binding.
 SQ SEQUENCE 553 AA; 59939 MW; F2A60424416B8C3F CRC64;

Query Match 5.1%; Score 148; DB 2; Length 553;
 Best Local Similarity 20.6%; Pred. No. 0.0012;
 Matches 114; Conservative 58; Mismatches 222; Indels 160; Gaps 21;

QY 52 LDGCAAPSAGSSAIGKNGKPHOREFENLSAEAVTRGVGMSHTWCSTPRHPP---NE 108
 DB 81 LAPAPLYPPNNYV---NVTGPSAGSFQ---QGYLTVGGTTWHWAASCRHHPSPDFVMK 134
 QY 109 SLPGIGRP-KLSNDPAEDDKWENLYSEARLIGT-----STKEFDESIRHTLVLR 158
 DB 135 SKYGVGRDWPISYDEME---PW---YCEAEYEGVAGPSPDSQSPSRSPRPPMDVFP 188
 QY 159 SLODAY-----KDRQIRFP---LPLAC-----HRLKN 183
 DB 189 AHGDTYFASVYVPHGYNLVPIPOGRSTRPWEGRPVCCGNNNCOPICPGMYNGIHHIER 248
 QY 184 APEYVEVHSAENLFHSIYNDKOKLFTLLTNHRCR---LALTGGYKKGIGAAEVRNLL 240
 DB 249 AESKGVAVLAESVYKIDTDD-----NNRVTAVHWLDNQGASHKATG----- 290
 QY 241 ATRNPPSOLDYSIMAKVYVILASGAIGNPQILYNSGFSGLQVTPRNDSLIPN-----LGR-- 294
 DB 291 -----KAFALACNGIETPRLLQA-----ANKANPTGIANSDDMVGRNN 329
 QY 295 -----YITQPMAFCQIVLRQEFVDSVRDPPVGLPWKEAVAQHIKAPTDALPIP 345
 DB 330 MDHSGFHCSELTPEPVNLGRGPAQSSCMVGRDGAFRSEY---SANKMILNISRVPAT 386
 QY 346 FRDPEQVTPFTTEHPHWTQIHRDAFSYGAAGPEVDSRVI-----VDLRWFGATDPEA 399
 DB 387 -----KQALAKGLVKALDEEIRYSIHGVLSLSLEPLPD 423
 QY 400 NNLLVFQNDVQDYSMPQPTFRYPSTASNVARKMADMCEVASNLGGYLPSTPPQFMD 459
 DB 424 ENRLTLKTRKDPHGLACPDIDYVDGVYRKGAATAAEQHQHIGSLFNG-----KEFN 476
 QY 460 PGLAL-----HLAGTTRIGFKATTVADNNSLVWDFANLYVAGNGTIRTFGEPNPTLSMC 515
 DB 477 ITTALNANNHMGTTGSKADAVVDGNCRTFDHENLWLPFGGAIPASVSVNNTLSMAA 536
 QY 516 HAKSARSILNTLK 529
 DB 537 LGLKAHDIISLRMK 550

RESULT 14

Q9R226 PRELIMINARY; PRT; 722 AA.
 ID Q9R226
 AC Q9R226;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE GMC oxidoreductase.
 GN DRA0127.
 OS Deinococcus radiodurans.
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
 OC Deinococcaceae; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=R1;
 RX MEDLINE=20036896; PubMed=10567266;

RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Croesby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Uterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 radiodurans R1.";
 RL Science 286:1571-1577(1999).
 DR EMBL; AE001862; AAF12230.1; -;
 DR TIGR; DRA0127;
 DR InterPro; IPR001064; Crystallin.
 DR InterPro; IPR001199; Cyt_B5.
 DR InterPro; IPR000172; GMC_oxred.
 DR Pfam; PF00732; GMC_oxred; 1.
 DR PROSITE; PS00225; CRYSTALLIN_BETAGAMMA; 1.
 DR PROSITE; PS00191; CYTOCHROME_B5_1; 1.
 DR PROSITE; PS00624; GMC_OXRED_2; 1.
 DR Complete proteome.
 SQ SEQUENCE 722 AA; 75376 MW; 450DFICD1B7596F9 CRC64;

Query Match 5.0%; Score 144; DB 16; Length 722;
 Best Local Similarity 21.1%; Pred. No. 0.004;
 Matches 107; Conservative 68; Mismatches 196; Indels 136; Gaps 26;

QY 90 GGMSTHTCSTPRIHPMESLPGICRPLSN-----DPAEDDKWENLYSEARLIG 141
 DB 277 GGSVTWNSV---PPRDDI---RQWASEHGLSDVADPGVD---RHDAVLERM-G 323
 QY 142 TSTKEFDESIRHTLVLSLDAYKQRIFRPLPLACHLKNAPVEVHSAENLFHSIY 201
 DB 324 VSEQCSDHNGPHQ---RLVEGADKLGYTEVKA-----ALNLSPEH---YDADKAGHAGF 371
 QY 202 NDD---KOKKLT-----LLTNHRCRLALTGGYKKGIGAAEVRNLLATRNPS 247
 DB 372 GDOTGAKQGLTNTFLKDAFEAGARILVGTARQVRLVEDG-----RAAGVSAVTMGDETR 426
 QY 248 QLDYSIMAKVYVILASGAIGNPQILYNSGFSGLQVTPRNDSLIPNLGRYITEQPMACQIV 307
 DB 427 QIT--VRAPQVVVACGALETPLLRSIGG-----PAAGRYLRLHPAGLVAGI 473
 QY 308 LRQEFVDSVRDDPYG---LPWKEAVAQHIANKPTDALTPIPRDPE-----P 351
 DB 474 -----YGEDQRAMWGP-----POSGILKQFADHENGHGFIIEGVQYGP 511
 QY 352 QVTT---PPT---EEH-PWHTQIHRDAFSYGAAGPEVDSRVID-----LRWFGATDP-EA 399
 DB 512 ALMASGLPMTGGEAHRDLMSKFRHATFVSIYQDRGHGQVTVUDEGNVHTYALTDDDA 571
 QY 400 NNLLVFQNDVQDYSM-----PQPTFRYPSTASNVARKMADMCEVAS---NLGGYLP 451
 DB 572 RN---FRRCVTESSIRLHEAAGAEIIVALAPGVAPWRRGDDLEAFIGQAVPLGAGQTV 628
 QY 452 TSPQFMDPGLALHLAGTTRIGFKATTVADNNSLVWDFANLYVAGNGTIRTFGEPNPTL 511
 DB 629 FS-----AHQWGSARMGSDPQTSVADPDQLHDVPGVWIGTSAFPPTCSGVNPMV 678

RESULT 15
 Q98C76 PRELIMINARY; PRT; 523 AA.
 ID Q98C76
 AC Q98C76;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
 DE Hypothetical protein mlr5266.
 GN MLR5266.
 OS Rhizobium loti (Mesorhizobium loti).

```

search completed: October 28, 2003, 15:59:49
job time : 109 secs

```